



Universitat Politècnica de Catalunya

Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona

Departament de Teoria del Senyal i Comunicació

ASSIGNACIÓ DE POTÈNCIA DE BAIXA COMPLEXITAT PER SISTEMES OFDMA-MULTIANTENA

Projecte final de carrera

Pol Henarejos Hernández

pol@redyc.com

Direcció: Ana I. Pérez-Neira

anuska@gps.tsc.upc.edu

* * *

Amb el suport de:

Centre Tecnològic de Telecomunicació de Catalunya - CTTC

NEWCOM++

Seventh Framework Programme

* * *

Barcelona - maig, 2009

“Si no mires de fer quelcom més enllà del que ja has aconseguit, mai creixeràs.”



Resum

Aquest projecte està destinat a tota aquella gent que senti curiositat per les telecomunicacions. Per llegir aquest projecte no es necessita cap coneixement previ. Poc a poc es van introduint els conceptes que fonamenten la telecomunicació. En ell, el lector hi pot trobar una recerca per tal de millorar els sistemes actuals multiusuari. Concretament, es centra en els escenaris de radiodifusió amb una estació base i terminals mòbils que utilitzen la modulació OFDM, com entorns WiMAX [WIM29], UMTS o per satèl·lit. A més, també s'ocupa d'incorporar la diversitat espacial mitjançant múltiples antenes en el transmissor. Així, tracta d'unir la diversitat freqüencial que presenta OFDMA amb la diversitat espacial que ofereix la conformació oportunista amb un algoritme de baixa complexitat. Addicionalment, basant-se amb el concepte de *cross-laying*, també es presenta com un mecanisme d'accés tant en freqüència com en espai.

La recerca posa els seus esforços en dos pilars que la conformen: 1) L'optimització de la potència per tal d'obtenir una velocitat global màxima i 2) la baixa complexitat. En altres paraules, hom trobarà una nova forma de maximitzar la velocitat total del sistema amb poca complexitat. A més a més, també s'hi incorporen prioritats en les transmissions per tal d'atorgar una porció determinada als usuaris de la velocitat total. Això permet incorporar el concepte de qualitat de servei.

Finalment, aquesta recerca està referendada per resultats obtinguts de varis simuladors. En aquests resultats es pot apreciar l'evolució que experimenta la velocitat global del sistema juntament amb la velocitat dels usuaris.

Tot plegat fa que aquest projecte mostri una nova perspectiva d'afrontar l'actual demanda de prestacions i serveis que es generen en les múltiples comunicacions.

D'aquest projecte han sortit algunes publicacions presentades en els congressos *International ITG Workshop on Smart Antennas 2009* celebrat a Berlín i *IEEE Globecom 2009 Wireless Communications Symposium*, encara pendent d'aprovació. També n'ha sortit una publicació de revista, presentada a *IEEE Transactions*, també pendent d'acceptació. La recerca compta amb el suport de la Comissió Europea sota el projecte NEWCOM++ (216715), Optimix (Grant Agreement 214625) i el Govern Espanyol TEC2008-06327-C03-01.



Agraïments

Tot i ser les primeres línies que el lector pot llegir, són les últimes que he escrit. Amb elles vull reflectir el que ha suposat escriure aquest projecte. Han estat uns mesos de dur treball però que el resultat, que hom té davant dels ulls, ha pogut mostrar un nou horitzó a explorar. El món de la investigació certament requereix molta imaginació i alta capacitat de pensament que, malauradament, a vegades, quan més es necessiten, no afloren. Però la satisfacció que dóna quan una nova idea apareix és quelcom difícil d'expressar. Aquest projecte, en el fons, no deixa de ser una idea que s'ha anat madurant i sofisticant i que, a més, encara li queda molt camí.

En primer lloc, voldria agrair el suport i la direcció de l'Ana Pérez. Ja d'abans la gran capacitat de treballar conjuntament era digna d'apreciar i valorar. L'assessorament, l'orientació i la comunicació són eines clau que m'ha ofert i que han permès escriure aquestes línies. També s'ha d'agrair l'entrega que li ha dedicat i, sobretot, l'oferta d'aquest projecte per desenvolupar-lo i les seves perspectives de futur. Gràcies de debò.

En segon lloc, també m'agradaria agrair la col·laboració del Velio Tralli en tota la recerca. La seva dedicació i esforços han estat clau per resoldre els problemes i conflictes que anaven apareixent. Tot i no haver treballat de manera conjunta anteriorment, s'ha bolcat amb una predisposició i comunicació que facilitaven molt el treball.

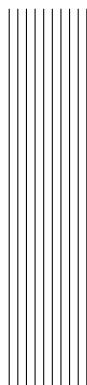
Com no, vull anomenar els amics que també saben els maldecaps i alegries que aquesta feina comporta. A la Maria, a l'Oriol, al Jordi, a la Sílvia. Ells comprenen què ha suposat escriure'l (perquè també han hagut de passar pel mateix).

Finalment, també guardo un agraïment especial als meus pares pel suport moral i incondicional. Dubtes i dilemes, no només tècnics, s'han fet menys importants amb la seva opinió. També, sobretot, a la Sara. El seu amor i la seva estima han estat un gran suport. I també, als seus pares per tota l'ajuda que m'han entregat. A tots ells, els vull agrair la seva paciència, perseverança i ajuts, tant en el redactat com en la lectura. Segurament ells estaven més interessats en que l'acabés d'escriure que jo mateix. Gràcies a tots.

I a tota la gent que s'ha interessat pel meu treball, que em preguntava com m'anava, quant em quedava per acabar i que per què servia. Sempre els he tingut presents. Voldria que tots ells poguessin arribar a familiaritzar-se amb el que faig, amb el que penso, amb el que sóc.

Gràcies, moltes gràcies.

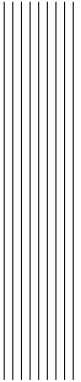
Pol. Barcelona, abril de 2009.



Índex

Notació i simbologia	xi
0.1 Notació	xi
0.2 Símbols	xii
Anglicismes	xv
Acrònims	xvii
1 Introducció	1
1.1 Motivació	1
1.2 Estructura del projecte	2
2 Fonaments de la telecomunicació	5
2.1 Orígens i història	5
2.2 Teoria de la informació	6
2.3 Comunicacions	8
2.4 Transmissió ràdio	10
2.5 Paràmetres qualitatius	11
2.5.1 SNR	11
2.5.2 Capacitat	12
2.5.3 BER	12
2.6 Canal de radiocomunicació	12
2.7 Diversitat freqüencial	15
2.7.1 OFDM i OFDMA	15
2.8 Diversitat espacial	16
2.8.1 Capacitat	19
3 Sistemes MIMO Multiusuari	23
3.1 Accés múltiple	24
3.2 Radiodifusió	26
3.2.1 Estratègia òptima	27
3.2.2 Implementacions òptimes	29
3.2.3 Implementacions subòptimes	31

3.2.4	Regió QoS	36
4	Assignació de potència per OFDMA-multiantena	37
4.1	Escenari	37
4.2	Optimització dual	39
4.3	Algoritme	42
4.3.1	Planificador espacial	42
4.3.2	Planificador freqüencial	44
4.4	Convergència	45
4.5	Estabilitat	46
4.6	Complexitat	47
4.6.1	Complexitat reduïda	48
4.7	<i>Feedback</i>	48
4.8	<i>Fairness</i>	49
4.9	Resum	50
5	Resultats	53
5.1	Escenari senzill	53
5.1.1	Actuació	53
5.1.2	Comparativa	58
5.2	Escenari real	63
6	Conclusions i treball futur	67
A	Apèndix A. Apèndix matemàtic	71
A.1	Teoria matricial	71
A.1.1	Autovalors i autovectors	71
A.1.2	Descomposició en valors singulars. SVD	71
A.2	Propietats emprades	72
A.3	Derivada de la funció objectiu	72
A.4	Funcions de probabilitat	73
A.4.1	Rayleigh	73
A.4.2	Rice	73
A.4.3	Lognormal	74
A.5	Optimització dual	76
A.6	Ergodicitat	79
B	Apèndix B. Publicacions	81
	Índex de figures	98
	Índex de taules	99
	Bibliografia	101



Notació i simbologia

Degut a que no hi ha una notació estàndard per aquesta temàtica, aquest projecte ha emprat la que a continuació es detalla. Durant tot el projecte s'utilitza sempre la mateixa.

0.1 Notació

- a Escalar, paràmetre, constant.
- \mathbf{a} Vector, en columna.
- \mathbf{A} Matriu.
- \mathbf{a}^T Vector \mathbf{a} transposat. Equival a un vector en fila.
- \mathbf{a}^H Vector \mathbf{a} hermític. Equival a un vector en fila amb les components conjugades.
- \mathbf{A}^T Matriu \mathbf{A} transposada.
- \mathbf{A}^* Matriu \mathbf{A} conjugada. Equival a conjugar cada element.
- \mathbf{A}^H Matriu \mathbf{A} hermítica. Equival a $(\mathbf{A}^T)^*$.
- $\sqrt{\mathbf{A}}$ Arrel quadrada d'una matriu. És una altra matriu tal que $\sqrt{\mathbf{A}}\sqrt{\mathbf{A}} = \mathbf{A}$. No s'ha de confondre en realitzar l'arrel quadrada de cada element.
- $\frac{1}{\mathbf{A}}$ Matriu on cada element és $\frac{1}{a_{ij}}$.
- a_{ij} Element de la fila i i de la columna j de la matriu \mathbf{A} .
- $|a|$ Mòdul de l'escalar complex a .

$\ \mathbf{a}\ $	Norma Euclideana del vector \mathbf{a} .
$x(t) * y(t)$	Convolució de les funcions $x(t)$ i $y(t)$.
$\mathbf{A} \otimes \mathbf{B}$	Producte de Kronecker.
$\mathbb{E}\{\mathcal{X}\}$	Esperança matemàtica de la variable aleatòria \mathcal{X} . Equival a $\int_{-\infty}^{+\infty} x f_{\mathcal{X}}(x) dx$.
$\text{tr}(\mathbf{A})$	Traça de la matriu \mathbf{A} . Equival a $\sum_i a_{ii}$.
$\text{diag}(a_1 \dots a_M)$	Matriu diagonal on els elements de la diagonal són a_i .
$\det(\mathbf{A}), \mathbf{A} $	Determinant de la matriu \mathbf{A} .
$\text{rk}(\mathbf{A})$	Rang de la matriu \mathbf{A} .
\mathbb{R}	Cos dels nombres reals.
\mathbb{R}^+	Cos dels nombres reals positius (el zero no està inclòs). Equival al conjunt $(0, +\infty)$.
\mathbb{R}^n	Vectors de n components reals.
$\mathbb{R}^{n \times m}$	Matrius de n files i m columnes d'elements reals.
\mathbb{C}	Cos dels nombres complexos.
\mathbb{C}^n	Vectors de n components complexos.
$\mathbb{C}^{n \times m}$	Matrius de n files i m columnes d'elements complexos.
$\arg \max_x f(x)$	El valor de x que maximitza la funció $f(x)$.
$\text{erf}(x)$	Funció d'error.

0.2 Símbols

K	Nombre d'usuaris.
k	Índex d'usuari.
M	Nombre de portadores.

m	Índex de portadora.
N_T	Nombre d'antenes en transmissió.
N_R	Nombre d'antenes en recepció.
q	Índex de <i>beam</i> .
j	Índex de permutació.
L	Nombre de camins.
l	Índex de camí.
N_0	Potència del soroll.
σ_w^2	Potència del soroll. Equivalent a N_0 .
γ	SINR instantània.
ξ	CINR instantània.
\mathbf{I}_m	Matriu identitat de rang m .
$\mathbf{1}_m$	Vector de tot 1 de dimensió m .
\mathbf{h}, \mathbf{H}	Vector, matriu de canal.
C	Capacitat, en <i>bps</i> .
P	Potència transmesa.
\bar{P}	Potència mitja total disponible en el sistema.
P_s	Potència de sondeig.
Q	Nombre de <i>beams</i> .
N_M	Rang de la matriu $\tilde{\mathbf{H}}$.
x, \mathbf{x}	Senyal, vector de transmissió.
s, \mathbf{s}	Símbol, vector de símbols.
w, \mathbf{w}	Soroll, vector de soroll.

r, \mathbf{r} Senyal, vector de senyal rebut.

y, \mathbf{y} Senyal, vector de senyal processat.



Anglicismes

La presència global de les telecomunicacions fa que s'utilitzi l'anglès com a llengua vehicular. Per aquesta raó molts dels mots i expressions emprats no tenen una traducció literal al català. D'altres, sí la tenen però per facilitat d'expressió i d'escriptura s'ometen i s'utilitzen en anglès.

Aquest projecte ha pretès utilitzar les traduccions tant com ha estat possible. Els termes no traduïts s'han escrit en cursiva per facilitar la lectura. A continuació s'escriuen els anglicismes emprats amb una breu accepció a manera de traducció.

array	Conjunt d'antenes.
beam	Feix.
beamforming	Conformació de feix.
benchmark	Cota de referència.
broadcasting	Radiodifusió. Una estació base transmet a múltiples usuaris.
cross-laying	Creuament de capes. Correspon a la interacció entre les capes adjacents PHY i MAC.
downlink	Enllaç de descàrrega. Direcció de l'estació base a l'usuari.
duality gap	Error dual. Diferència entre el màxim del problema inicial i mínim del problema dual.
feedback	Realimentació. Informació de suport de l'usuari cap a l'estació base.
outage	Probabilitat que un paràmetre sigui superior a un llindar.
path-loss	Pèrdues per distància.
time-sharing	Compartició del temps.

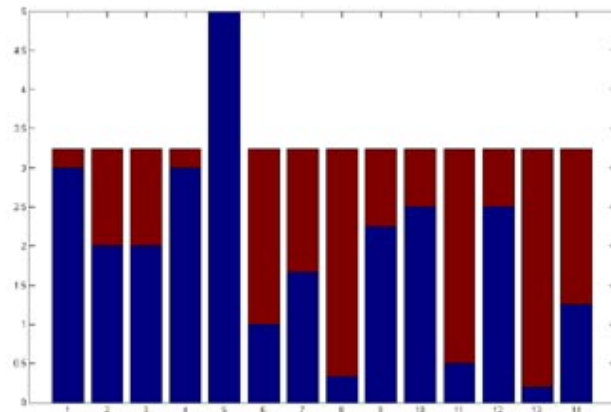


Figura 1: Interpretació visual del *water-filling*. La potència assignada és l'àrea en vermell per sota d'un llindar.

s.t.	<i>Subject to</i> . Subjecte a. Limitat per.
shadowing	Efecte d'ombra. Disminució del nivell de potència rebut per zones d'ombra.
slot	Porció o cel·la.
steering	Vector de direccions.
sumrate	Velocitat global del sistema. Si no s'especifica el contrari es suposa el valor instantani d'aquesta.
uplink	Enllaç de càrrega. Direcció de l'usuari a l'estació base.
water-filling	Tècnica per assignar potència. La potència assignada és l'àrea per sota d'un llindar. Veure figura 1.



Acrònims

En el projecte s'han utilitzat els següents acrònims i sigles. Cal afegir que la gran majoria d'ells provenen de l'anglès i s'han mantingut sense traducció.

3GPP	<i>Third Generation Partnership.</i>
AM	<i>Amplitud Modulada.</i>
ATT	<i>American Telephone and Telegraph.</i>
AWGN	<i>Additive White Gaussian Noise.</i>
BC	<i>Broadcast.</i>
BER	<i>Bit Error Rate.</i>
CINR	<i>Channel to Interference plus Noise power Ratio.</i>
CSI	<i>Channel State Information.</i>
CSIR	<i>CSI at the Receiver.</i>
CSIT	<i>CSI at the Transmitter.</i>
DL	<i>Downlink.</i>
DPC	<i>Dirty Paper Coding.</i>
ETSI	<i>European Telecommunications Standards Institute.</i>
FFT	<i>Fast Fourier Transform.</i>
FIR	<i>Finite Impulse Response.</i>

FM	<i>Freqüència Modulada.</i>
HSDPA	<i>High Speed Downlink Packet Access.</i>
ITU	<i>International Telecommunication Union.</i>
LOS	<i>Line Of Sight.</i>
LTE	<i>Long Term Evolution.</i>
MAC	<i>Multiple Access.</i>
MIMO	<i>Multiple Input Multiple Output.</i>
MISO	<i>Multiple Input Single Output.</i>
MOB	<i>Multiuser Opportunistic Beamforming.</i>
NLOS	<i>Near Line Of Sight.</i>
OFDM	<i>Orthogonal Frequency Division Multiplexing.</i>
OFDMA	<i>Orthogonal Frequency Division Multiple Access.</i>
OSI	<i>Open Systems Interconnection.</i>
PHY	<i>Physical.</i>
QoS	<i>Quality of Service.</i>
SINR	<i>Signal Interference Noise Ratio.</i>
SIMO	<i>Single Input Multiple Output.</i>
SISO	<i>Single Input Single Output.</i>
SNR	<i>Signal to Noise power Ratio.</i>
SVD	<i>Singular Value Decomposition.</i>
TDD	<i>Time Division Duplex.</i>
TDMA	<i>Time Division Multiple Access.</i>
TIC	<i>Tecnologies de la Informació i Comunicació.</i>

UPA	<i>Uniform Power Allocation.</i>
UL	<i>Uplink.</i>
ULA	<i>Uniform Linear Array.</i>
UMTS	<i>Universal Mobile Telecommunication System.</i>
WiMAX	<i>Worldwide Interoperability for Microwave Access.</i>
ZF	<i>Zero Forcer.</i>



1 Introducció

Aquest projecte presenta una nova tècnica de gestió de recursos en escenaris de *broadcasting*. Concretament gestiona l'accés dels usuaris mitjançant una assignació de potència utilitzant una implementació de baixa complexitat. A més, també se'ls assigna una prioritat en la velocitat per tal d'assolir una determinada QoS. Aquest projecte té una fàcil implementació en sistemes comercials ja coneguts que incorporin OFDMA.

1.1 Motivació

L'àmbit de les telecomunicacions ha sofert una revolució els últims anys gràcies a l'avenç de la tecnologia, tant per ús militar, industrial, acadèmic o civil. L'aparició de la telefonia mòbil o l'Internet sense cables han fet que la demanda de noves infraestructures i millors prestacions augmentés estrepitosament. A més, el creixement i la subscripció de cada cop més usuaris ha fet que la situació actual quedés limitada.

Tots aquests factors han desencadenat que la investigació actual es centri, sobretot, en oferir més i millors prestacions sense dedicar més esforç addicional. Aquest esforç queda reflectit, primordialment, en l'ús o el consum d'energia. El consum d'energia pren una important rellevància perquè es tradueix en una despesa. L'ús d'energia costa diners i un consum elevat ja es pot preveure que no serà eficient, per moltes prestacions que ofereixi. Per aquesta raó, sempre es crea un compromís entre oferiment i cost. És a dir, quin cost té oferir certes prestacions.

Una de les noves vies que s'han obert és la recerca de nous tipus de diversitat. La diversitat és la manera de transmetre informació de múltiples formes. Per exemple, transmetre una paraula en diferents intervals de temps farà que es rebi múltiples vegades i que el receptor la pugui processar més correctament. Dit d'altra manera, es tracta de fer que el receptor tingui més informació del que s'ha transmès. Això permet realitzar múltiples aplicacions, com la supressió d'errors o d'interferències.

De manera particular, la diversitat espacial ha obert noves tècniques de processament de la informació. Per obtenir aquesta millora s'utilitzen múltiples antenes en transmissió o recepció per tal de fer front als esvaïments del canal, per transmetre informació paral·lelament o per transmetre de forma direccional. Aquest ús queda ben sintetitzat en les tècniques MIMO.

D'altra banda, la diversitat freqüencial permet utilitzar múltiples freqüències per transmetre informació de forma paral·lela per modulacions i filtrats a diferents freqüències. A més, la diversitat temporal també permet transmetre aquesta mateixa informació en instants dife-

rents.

Aquest projecte pretén mostrar un sistema de comunicació nou i original fent ús de les diversitats anteriors per tal d'augmentar la velocitat de les transmissions. L'escenari que es contempla és un escenari de radiodifusió o *broadcasting*. Aquests escenaris consisteixen en un transmissor que transmet informació a múltiples usuaris. Són particularment atractius perquè l'augment de les TIC afavoreix l'aparició de més escenaris com aquests. Exemples en són la televisió o la comunicació de les estacions base amb els terminals mòbils.

En altres paraules, aquest projecte proposa una nova alternativa per satisfer a més usuaris amb més velocitat utilitzant tècniques i mecanismes d'accés actuals, com és OFDMA. Òbviament, tot això respectant sempre un consum limitat de potència.

A més a més, també s'introdueix el concepte de prioritats o pesos per usuari. Aquestes prioritats són uns coeficients que limiten la velocitat de l'usuari proporcionalment. Ofereixen un percentatge de la velocitat total de la que disposa el sistema a l'usuari per evitar monopolis i realitzar un ús òptim d'aquesta velocitat. En altres paraules, reparteix el pastís entre els usuaris assignant porcions diferents segons les seves necessitats. Resulta evident que un usuari que descarregui veu no necessitarà tanta velocitat que un que estigui veient una pel·lícula en temps real. Per què s'ha de destinar recursos a usuaris que en aquell moment no els utilitzen? Aquest reús dels recursos fa que la resposta global del sistema sigui més eficient.

Una altra de les intencions d'aquest projecte és facilitar la seva lectura i comprensió. Per això té com a objectiu que no siguin necessaris coneixements previs de comunicació. Senzillament amb la lectura ha de resultar suficient per arribar entendre tot el procediment. No obstant, sí es donen per apresos conceptes i notacions matemàtiques.

1.2 Estructura del projecte

Aquest projecte està estructurat en diferents capítols de temàtica diferent. La figura 1.1 reprodueix l'esquema que s'ha seguit.

El **capítol 2** introdueix els conceptes de la telecomunicació per comprendre tot el projecte. Òbviament l'explicació resulta molt sintetitzada. Aspectes com la teoria de la informació —essencial pel plantejament d'aquest projecte—, sistemes de comunicació, paràmetres qualitatius, el canal de radiocomunicació o els tipus de diversitat són presentats de manera ordenada. Un cop finalitzada la seva lectura, el lector hauria de tenir present tots aquests conceptes dels quals es derivarà la idea posterior.

El **capítol 3** dedica un punt sencer als sistemes MIMO multiusuari. En ell, es planteja tota la problemàtica que es deriva quan s'introdueixen múltiples usuaris. Paràmetres com la capacitat queden alterats i són requerides altres tècniques. Per tant, un cop finalitzi el capítol, juntament amb els coneixements del capítol anterior, el lector hauria d'estar preparat per comprendre la recerca duta a terme.

El **capítol 4** esdevé el nucli d'aquest projecte. En ell, es proposa l'objectiu que es vol assolir. Per fer-ho, primerament es planteja l'escenari que es contempla. Per això, el lector ha de tenir present tots els conceptes per tal d'arribar a comprendre el plantejament. Seguidament s'introdueix el concepte d'optimització dual per donar solució al problema amb dues restriccions. A continuació presenta l'algoritme proposat que implementa la solució al problema. Finalment s'analitzen aspectes com l'estabilitat, la complexitat, els paràmetres de *feedback* o la justícia del sistema.

El **capítol 5** mostra els resultats que s'obtenen al capítol anterior. Per fer-ho s'utilitzen múltiples gràfiques de diverses simulacions. Addicionalment també es comenten i s'analitzen

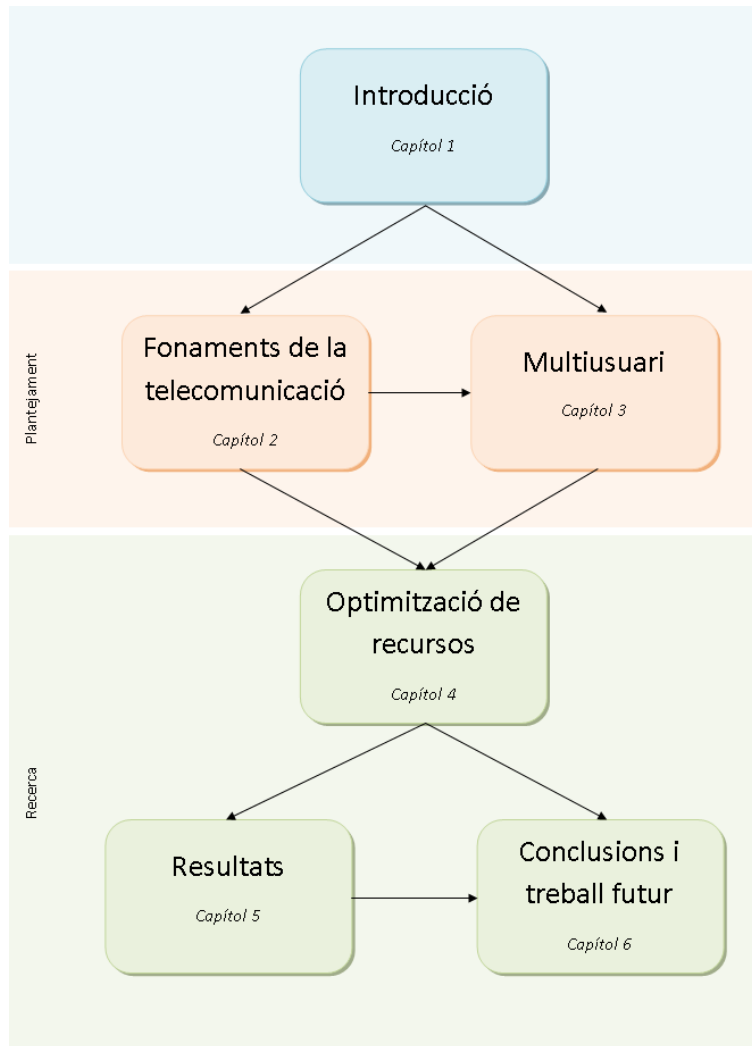


Figura 1.1: Esquema del projecte.

els resultats a partir de les explicacions del capítol previ.

Finalment, el **capítol 6** presenta la situació actual i les noves vies de recerca que s'obren amb aquest projecte.

En els **annexos** s'inclouen les consideracions matemàtiques que s'han utilitzat en aquest projecte i les publicacions de les que n'ha sortit el mateix.



2 Fonaments de la telecomunicació

2.1 Orígens i història

LA manera de comunicar-se sempre ha estat en constant evolució. Però no va ser fins a finals del segle XIX i principis segle XX que no va sorgir la revolució de les telecomunicacions. El descobriment de les equacions de Maxwell [MAX01] el març de 1861 va permetre l'ús i manipulació dels camps electromagnètics per transportar energia en forma d'informació.

A principis del segle XX es van iniciar les comunicacions mòbils. En un principi s'establiren entre el transport marítim i el terrestre. Així, els vaixells podien comunicar-se mitjançant tons amb altres vaixells o amb les estacions que estaven a terra ferma. Però no va ser fins als inicis dels anys 20 que la policia de Detroit i Nova York es van posar d'acord per crear un sistema de missatgeria. Aquest consistia en un conjunt de tons d'avís. D'aquesta manera es començaven a unir grups de terminals en ciutats diferents d'una certa distància.

Posteriorment, cap a finals dels anys 20 la policia de Cleveland va començar a utilitzar les modulacions coherents. Concretament, es van crear els primers receptors superheterodins per les modulacions en AM. No obstant, fins els anys 30 no van ser comunicacions bidireccionals. Fins aleshores, el canal de transmissió només anava en una direcció.

L'any 1932 la Unió Internacional de Telegrafia, fundada anys enrere, es va unir amb la Unió Radiotelegràfica Internacional per fundar la Unió Internacional de Telecomunicacions (ITU). No obstant, no va ser fins l'any 1934 que prengué aquest nom. Aquesta entitat regulà des d'aleshores els estàndards universals, mecanismes de control i, sobretot, la gestió de l'espectre radioelèctric.

Després de l'aparició de l'AM, cap al final de la dècada dels 30 va crear-se la modulació en FM. Això va suposar un gran avenç pel gran nombre d'aplicacions que se li van donar. L'any 1946 l'empresa de telecomunicacions americana ATT va establir el primer servei de radiotelefonía mòbil per l'ús civil. Tot i ser terminals molt voluminosos i de poca bateria, el gran públic ja podia comunicar-se a distància. En aquells anys, l'encaminament entre les trucades encara era manual i es feia mitjançant una operadora. Fins el 1964 no va ser automàtic i amb un assignació dinàmica.

Poc a poc van anar desenvolupant-se i anà sorgint una teoria que donaria cos i forma a la informació. L'any 1948, Claude E. Shannon va publicar l'article *Una teoria matemàtica de la comunicació* [SHA02] que establiria els fonaments matemàtics de la telecomunicació. Aquest article se'l coneix com a la *Teoria de la informació*.

Anys més tard, cap a la dècada dels 60, es va començar a popularitzar el model de capes OSI. Aquesta arquitectura es va instaurar com un model a tenir present a l'hora de dissenyar aspectes d'un sistema de comunicació, des del nivell físic fins al nivell més elevat com la gestió de la xarxa.

Ja en plena dècada dels 70, Gallager va establir una teoria que unia el que en aquell moment eren models disperss: la teoria de la informació i la teoria de xarxes. D'aquesta manera es podia aplicar la teoria per dissenyar xarxes. Van començar a sorgir els primers protocols. L'any 1977 Gallager va resoldre un problema d'encaminament mitjançant l'optimització convexa.

L'any 1988 la Conferència d'Administracions de Correus i Telecomunicacions van crear l'Institut Europeu de Normes de Telecomunicacions (ETSI). Aquest organisme regula la normativa de telecomunicacions dins d'Europa.

Finalment, amb l'inici del segle XXI, les telecomunicacions van experimentar una revolució causada principalment per l'aparició amb força de l'Internet i la telefonia mòbil 2G. Des d'aleshores, les diferents àrees han sofert un creixement exponencial que encara es manté a dia d'avui.

2.2 Teoria de la informació

El concepte d'informació resulta ser sovint força ambigu i difús. No obstant, és possible definir una magnitud d'informació. Shannon va concebre una teoria que donaria forma i consistència a la informació i com es vertebra. El que va escriure va donar peu a una immensa varietat d'aplicacions. Gràcies a aquest punt d'inflexió, les telecomunicacions van desenvolupar-se fins al que es coneix avui.

La *Teoria de la informació* defineix el concepte d'informació, les seves propietats i els seus paràmetres, entre d'altres aspectes. Primerament, el que es necessita és trobar una manera de quantificar la magnitud de la informació. És a dir, poder definir una quantitat d'informació, conèixer si determinades situacions aporten o no aporten informació. Per això, s'utilitza la probabilitat d'aquests fenòmens.

Així doncs, donat un succés X , la informació que proporciona és major quant menys probable resulta. És a dir, "*La informació és funció inversa de la probabilitat.*"

$$\mathcal{I}(X) = \log_2 \left(\frac{1}{p(X)} \right). \quad (2.1)$$

La funció logarítmica compleix les propietats de continuïtat, monotonia creixent i, a més, es té que $\mathcal{I}(X_1 \cap X_2) = \mathcal{I}(X_1) + \mathcal{I}(X_2)$ sempre que els successos siguin independents. La base del logaritme defineix les unitats de la informació. Correspon al nombre d'estats estables que compon aquesta informació. Habitualment s'empra la base 2 perquè es relaciona amb els estats *actiu* i *no actiu* que pot tenir un circuit *flip-flop*. En aquest cas, les unitats de la informació s'anomenen *bits per símbol*. Els símbols s'atribueixen a la informació qualitativa i els bits a com són codificats aquests símbols. Per exemple, per un dau, els símbols correspondrien als 6 possibles valors que marquen cada una de les cares. Els bits codificarien aquests 6 símbols només amb dos estats (habitualment 0 i 1). Per tant, només serien necessaris 3 bits per codificar els 6 símbols. Així doncs, es posa de manifest la relació amb el nombre de bits b i el nombre de símbols M

$$M = 2^b \text{ o bé } b = \log_2 M. \quad (2.2)$$

Noti's que existeixen múltiples codificacions possibles.

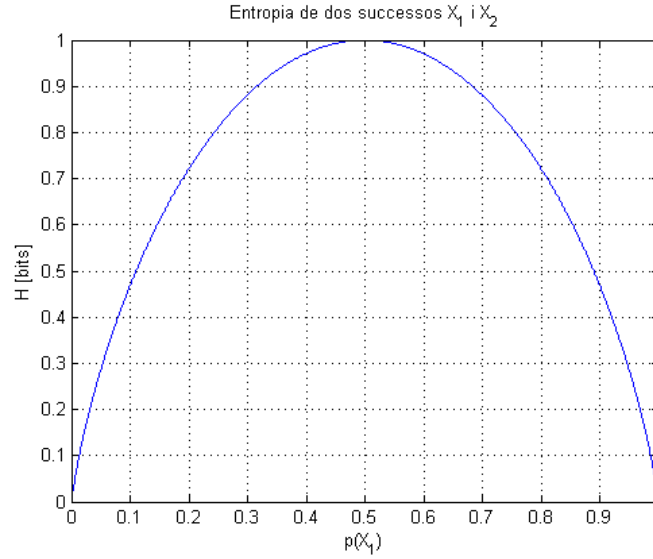


Figura 2.1: Entropia de dos successos amb probabilitats $p(X_1), p(X_2) = 1 - p(X_1)$

Anàlogament, donat un conjunt de N successos X_1, \dots, X_N tal que $\sum_i^N p(X_i) = 1$, es defineix la funció *entropia* com

$$\mathcal{H}(X_1, \dots, X_N) = - \sum_i p(X_i) \log_2(p(X_i)) \quad (2.3)$$

i mostra la incertesa d'aquest conjunt. La màxima incertesa es té quan l'entropia és màxima. En aquest cas, $\mathcal{H} = 1$ i es produeix quan $p(X_1) = \dots = p(X_N) = \frac{1}{N}$, és a dir, quan tots els successos són equiprobables. Que tots els fenòmens siguin equiprobables manté la idea d'incertesa absoluta. D'altra banda, l'entropia és mínima —i, per tant, la incertesa, també—, $\mathcal{H} = 0$ quan tots els successos són zero excepte un. Això indica que sempre es produeix el mateix succés i, per tant, la incertesa és nul·la. La figura 2.1 mostra l'entropia de dos successos.

Suposant que es disposen de dos conjunts, també es pot definir l'entropia conjunta com

$$\mathcal{H}(X, Y) = - \sum_{i,j} p(X_i, Y_j) \log_2 p(X_i, Y_j) = \mathcal{H}(X) + \mathcal{H}(Y|X) = \mathcal{H}(Y) + \mathcal{H}(X|Y) \quad (2.4)$$

on

$$\mathcal{H}(Y|X) = - \sum_{i,j} p(X_i, Y_j) \log_2 p(Y_j|X_i) \quad (2.5)$$

és l'entropia de Y condicionada a X .

Finalment, la informació mútua d'aquests successos ve donada per

$$\mathcal{I}(X, Y) = \mathcal{H}(X) - \mathcal{H}(Y|X) = \mathcal{H}(Y) - \mathcal{H}(X|Y) = \mathcal{H}(X) + \mathcal{H}(Y) - \mathcal{H}(X, Y). \quad (2.6)$$

Les unitats són *bits per segon* o *bps*. Així doncs, correspon a la quantitat d'informació (bits) que es poden intercanviar per unitat de temps (segons). És a dir, és la *velocitat de transmissió* o *rate*.

És fàcil d'observar que segons la codificació s'obtindrà una millor o pitjor velocitat de transmissió. A més, també es pot observar que, donat que el número de codificacions possibles

és finit, aquesta velocitat de transmissió assolirà un màxim per una codificació concreta. Així doncs, es pot definir la *capacitat* d'una transmissió com el màxim d'informació mútua:

$$\mathcal{C} = \max \mathcal{I}(X, Y). \quad (2.7)$$

Aquesta equació acaba de definir un llinar o límit de velocitat de transmissió. Aquesta magnitud resulta molt interessant perquè, tal i com es veurà més endavant, marca la rapidesa de transmissió que es pot aconseguir.

2.3 Comunicacions

A partir de la teoria de la informació exposada a l'anterior secció, es pot dissenyar un model de comunicacions. Anomenant X el conjunt de símbols que es vol transmetre d'una font i Y el conjunt de símbols rebuts en una altra font, es pot observar que la velocitat de transmissió màxima vindrà donada per 2.7. Així doncs, es pot interpretar com si el medi entre el transmissor X i el receptor Y tingués unes característiques concretes. És el que s'anomena *canal de comunicació*. Hom ho pot imaginar com si fos una canonada d'un diàmetre concret per la que només hi podrà anar un cabdal màxim determinat i no més.

El canal de comunicació és el medi físic per on es transmet la informació entre el transmissor i el receptor. Conté unes característiques pròpies (temps de resposta, transitori, potència, etc.) que limiten el transcurs de la comunicació i la seva velocitat. És a dir, és una aplicació matemàtica entre un cos d'origen i un cos de destí. Aquesta aplicació pot ser determinista i coneguda a priori o bé totalment aleatòria. A més, pot ser una aplicació lineal o no.

Hi ha nombroses formes de transmetre informació. Per això s'utilitzen els camps electromagnètics per contenir aquesta informació. És el que s'anomena *modulació*. La informació es pot modular en amplitud, fase, freqüència, polaritat, etc. del camp electromagnètic. Encara que el camp elèctric sempre té un camp magnètic associat i viceversa, s'acostuma a manipular el camp elèctric. Per tant, variant l'amplitud d'un camp elèctric, la seva fase o freqüència és possible transmetre certa informació. La funció que descriu aquest camp és el que s'anomena *senyal*.

Habitualment, els canals de telecomunicació solen modificar la informació tant en amplitud com en fase. Així doncs, a partir del model dels nombres complexos, equival a formular l'anterior com el producte del símbol, també complex, per un coeficient complex del canal. El resultat és un altre nombre de diferent amplitud i fase. Aquest resultat pot ser determinista o aleatori segons la naturalesa del canal.

Pel cas determinista, on el canal no és aleatori i produeix una sortida predictable, és possible “desfer” aquesta manipulació si es coneix el canal a priori. Aquesta operació senzillament consisteix en sondejar la resposta del canal davant del conjunt de símbols prèviament coneguts. S'intenta crear un mapa per obtenir una associació entre símbol d'entrada i de sortida. Un cop es té aquesta resposta, el receptor pot decidir equalitzar-la i obtenir el símbol transmès¹, segons l'estratègia que empli. No obstant, aquest procés pot requerir, segons l'escenari, potència addicional. Però coneixent la resposta del canal o la forma d'aquesta aplicació és possible detectar la informació transmesa.

El coneixement d'aquesta funció del canal, que s'anomenarà $h(t)$, implica un coneixement a priori per manipular la informació que pot ésser emprat en la codificació per fer front al

¹Cal tenir present que no s'obindrà el símbol transmès ja que contindrà soroll.

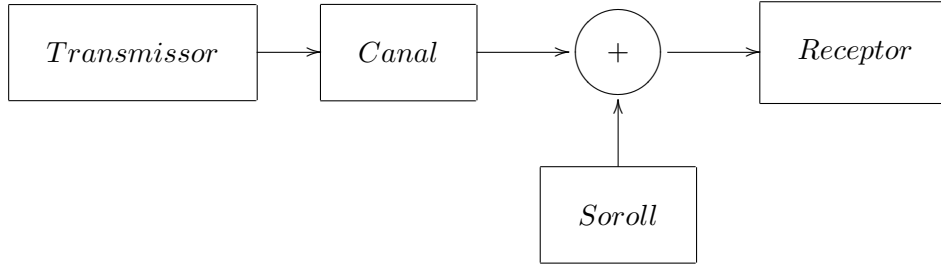


Figura 2.2: Sistema de comunicació

canal. Tenint en compte que es tracta d'un producte amb memòria, l'operació intrínseca que realitza el canal correspon a la convolució matemàtica.

A partir de la segona igualtat de 2.6 es pot apreciar com la informació rebuda depèn de la informació transmesa extraient la informació pròpia de Y tenint en compte el que s'ha rebut. En altres paraules, es pot modelar com si aquesta informació contingués pertorbacions causades per la recepció. És el que s'anomena *soroll* i no és més que la inexactitud que es produeix a la transmissió. Aquest soroll és aleatori i els valors que pot prendre són incerts. No obstant, aquesta pertorbació el que fa és desplaçar la informació rebuda amb més o menys potència. Per tant, es pot modelar com una variable aleatòria amb un valor esperat i una variància definides.

Definint w com la variable aleatòria del soroll, el seu valor esperat és la seva esperança matemàtica i la seva potència és la variància. Aquests paràmetres vénen donats per les expressions

$$\mu_w = \int_{-\infty}^{+\infty} x f_w(x) dx = \mathbb{E}\{w\} \quad (2.8)$$

$$\sigma_w^2 = \int_{-\infty}^{+\infty} (x - \mu_w)^2 f_w(x) dx = N_0. \quad (2.9)$$

S'anomena *soroll blanc* aquell que té nul el valor esperat i és incorrelat. És a dir,

$$\mu_w = 0 \quad (2.10)$$

$$R_w(k) = \mathbb{E}\{w(t)w(t + \Lambda)\} = \sigma_w^2 \delta(\Lambda). \quad (2.11)$$

En altres paraules, el soroll no guarda cap relació si es mesura en instants diferents.

D'altra banda, el *soroll acolorit* és aquell que la correlació no és nul·la.

Agrupant tot l'anterior, es pot sintetitzar tot el sistema de comunicació mitjançant la figura 2.2. Així doncs, la informació al receptor es pot escriure com

$$y(t) = h(t) * x(t) + w(t) \quad (2.12)$$

on l'operador $(*)$ denota la convolució entre els operands i equival a

$$h(t) * x(t) = \int_{-\infty}^{+\infty} h(t - \tau)x(\tau) d\tau. \quad (2.13)$$

Recuperant l'expressió de la capacitat 2.7, en un sistema de comunicació simple punt a punt deriva a

$$C = \log_2 \left(1 + \frac{P_y}{\sigma_w^2} \right) = \log_2 \left(1 + \frac{|y(t)|^2}{N_0} \right). \quad (2.14)$$

Això mostra que la capacitat del canal ve donada per la potència que es rep i per la influència de la potència del soroll.

2.4 Transmissió ràdio

Com s'ha vist a la secció anterior, els mitjans de comunicació poden ser tant variats segons la naturalesa del canal per on es transmetin. No és la intenció d'aquest projecte mostrar tots els canals de difusió d'informació. De forma particular, pren interès el canal de ràdio. S'anomena *canal de ràdio* aquell que no conté cap suport físic més que el propi medi. Així doncs, canals de ràdio són l'aire, l'aigua o el buit, entre d'altres. A diferència dels sistemes punt a punt, on la informació només es produeix entre un sol parell de transmissor-receptor, la radiodifusió contempla l'opció de transmetre a múltiples receptors o rebre de múltiples transmissors.

La radiodifusió ha cobrat una important rellevància en les últimes dècades gràcies a la televisió o a la telefonia mòbil. Això ha fet que es desenvolupés una tecnologia adient per cada situació però també ha fet que sorgissin nous problemes. La tendència actual és la d'oferir més serveis a més velocitat i a un número major d'usuaris a la vegada. Resulta essencial trobar noves eines o opcions per barallar o multiplexar la informació i alhora arribar a un número major d'usuaris. El medi de radiodifusió resulta altament interferent. Constantment es produeixen col·lisions o interferències.

Paral·lelament resulta obvi que per regular els accessos dels usuaris ha d'existir un gestor. Aquest gestor marca i dóna accés als usuaris de forma alternada segons diferents estratègies. Dins del model OSI [ZIM03], ocupa una subcapa de la capa de *data-link*. És l'anomenada *MAC layer* o *capa MAC*. Aquesta capa té la funció de regular l'accés al medi de l'usuari.

D'altra banda, també resulta primordial controlar els recursos físics que es dediquen a la comunicació. Això inclou, sobretot, la gestió i restricció de la potència. L'energia o potència que es dedica té efectes tant en termes de velocitat com econòmics. Per aquesta raó sol ser una restricció important a l'hora de dissenyar el sistema de comunicació. La gestió òptima de la potència constitueix un objectiu important dins de tota investigació o disseny. L'ús de potència delimita i constitueix el compromís entre prestacions i ús de recursos. Per aquest motiu les últimes investigacions focalitzen els seus resultats a trobar maneres cada cop més òptimes per brindar més prestacions sense haver d'augmentar els recursos necessàriament.

Aquesta gestió de potència es duu a terme en un nivell físic. La potència es regula modificant l'amplitud del camp electromagnètic que conté la informació a transmetre. Quanta més amplitud disposi més potència requerirà. Dins del model OSI, es situa a la capa física o *PHY layer*.

Així doncs, la interacció d'aquestes capes obre noves possibilitats ja que permet reutilitzar recursos d'una banda mitjançant les característiques de l'altra. Per exemple, un dels usos més interessants que se'n deriven és la gestió d'accés a través de la potència. Una forma de donar accés als usuaris és atorgant-los potència segons les necessitats del sistema. Els que no disposin de potència, no podran comunicar-se i, per tant, perden el seu accés. És el que s'anomena *cross-laying*.

Els tipus de radiodifusió es poden abordar des de tres punts de vista diferents. El primer sorgeix de les equacions de la Teoria de la Informació. Tal com s'ha narrat en el capítol anterior, els sistemes de comunicació es poden estudiar des d'una teoria matemàtica. Tot i ser força abstracta, aquesta teoria presenta els conceptes molt detalladament i resulta consistent a l'hora de resoldre situacions i problemes. A partir de les equacions més generals es poden anar particularitzant per l'escenari que es tracti i donar una resposta a priori del resultat. D'altra banda, aquest aspecte es pot analitzar des del Processament del Senyal. Aquest àmbit enfoca els seus esforços cap el disseny de sistemes i equips de radiocomunicació. Fa front a la problemàtica plantejant transmissors i receptors multiusuari que responguin amb les necessitats que es demanen. Finalment, tot l'anterior es pot estudiar a partir de la Teoria de Xarxes. Aquesta teoria és un altre mètode que es basa en altres models per trobar també respostes a les situacions que es plantegen.

Aquest projecte presenta els seus resultats basant-se sempre en la Teoria de la Informació per, tal com s'ha indicat, presentar els conceptes de forma més clara i concisa.

2.5 Paràmetres qualitatius

Els sistemes de comunicació poden ser més o menys fiables segons el seu disseny, la naturalesa del medi o per la precisió tecnològica amb la que estan fets. Per aquest motiu cal definir eines o paràmetres que aportin informació sobre la qualitat de la comunicació. Aquests orienten i indiquen quin grau de qualitat es té durant la comunicació i quines prestacions es poden arribar a aconseguir. Tal i com es veurà, aquests paràmetres es poden definir segons l'objectiu que es persegueix en termes de velocitat de transmissió, taxa d'error durant la detecció o qualitat del canal.

2.5.1 SNR

La SNR és la *Relació senyal a soroll* i indica la proporció de potència de senyal i soroll que es percep en una comunicació. És equivalent a dividir la quantitat de potència de senyal pròpia i desitjable amb la quantitat de potència del soroll a la recepció. Pel sistema bàsic de 2.12 la SNR serà

$$SNR = \frac{|h(t)|^2 P_x}{N_0}. \quad (2.15)$$

La SNR serà millor quanta més potència es transmeti, més potent sigui el canal o menys soroll es tingui.

Paral·lelament pels casos anteriors on apareixien interferències, aquestes també són considerades senyals no desitjats, juntament amb el soroll. Es defineix la SINR com la *Relació senyal a interferència i soroll*. Pel sistema plantejat a 3.11, la SINR d'un usuari k arbitrari serà

$$SINR_k = \frac{P_k}{\sum_{k' \neq k}^K P_{k'} + N_0}. \quad (2.16)$$

2.5.2 Capacitat

A partir de la teoria de la informació desenvolupada anteriorment, la capacitat mostra la velocitat de transmissió que es pot aconseguir. Una capacitat més alta indica que la qualitat de la comunicació millora. No obstant, no diu si és per una bona transmissió, detecció o per un baix nivell de soroll.

Hom pot percebre fàcilment que mesclant 2.15 i 2.14 s'obté una de les equacions bàsiques de la capacitat d'un sistema de telecomunicació

$$C = \log_2 (1 + SNR). \quad (2.17)$$

Més endavant es recuperarà aquest concepte per altres aplicacions.

2.5.3 BER

La BER correspon a la *taxa d'error per bit* i mostra la relació d'error que es produeix durant la detecció d'un bit. Dóna una idea del grau d'equivocació que comet el receptor. Es fa difícil trobar una expressió tancada que ofereixi una BER pel sistema, estratègia o escenari tractat. Depèn de factors com la modulació emprada, la codificació, l'entorn o d'altres. No obstant, una manera molt intuïtiva de valorar la BER a posteriori és transmetre una seqüència de bits, coneguda pel receptor, i realitzar l'operació

$$\frac{\text{nombre bits erronis}}{\text{nombre bits total}} \quad (2.18)$$

per obtenir una taxa instantània d'error. Donat que solen obtenir-se nombres força baixos, és habitual expressar l'anterior magnitud en decibels².

2.6 Canal de radiocomunicació

A la secció anterior s'ha vist quina importància té el canal durant la comunicació. En general, el canal contribueix directament a la potència multiplicant el senyal transmès per un coeficient intrínsec del canal. De canals de comunicació n'hi ha molts i de molt variats. Però pel cas que ocupa aquest projecte, el canal que s'estudia és el corresponent al canal de ràdio.

El canal de radiocomunicació segueix funcions variables amb múltiples components que cal comentar. El primer aspecte que compon el canal de radiocomunicació és les pèrdues de propagació. Durant la transmissió del camp electromagnètic, la potència es va propagant pel medi com si d'una bombolla es tractés. Això fa que en recepció es rebi menys potència. Resulta evident que quant més lluny estigui el receptor del transmissor, menys potència es percebrà. Aquesta atenuació ve donada pels factors ambientals, l'holografia de l'entorn o per la freqüència emprada. Així doncs, cada freqüència tindrà una atenuació diferent. En el pla temporal, es tradueix en un decreixement de la potència rebuda en funció de la distància.

D'altra banda, l'holografia del terreny fa que les ones electromagnètiques rebotin i canviïn de direcció. En entorns abruptes o amb molt mobiliari, com ambients urbans, es produiran moltes reflexions, anomenades *ecos*. El receptor rebrà múltiples propagacions de la mateixa informació rebuda. Aquestes propagacions seguiran distàncies diferents i, per tant, arribaran en moments diferents. Això causa modificacions de la fase i l'amplitud perquè aleatòriament es poden unir constructivament o destructivament. Per tant, en moments concrets es produiran

² $x|_{dB} = 10 \log_{10} (x|_{lineal})$

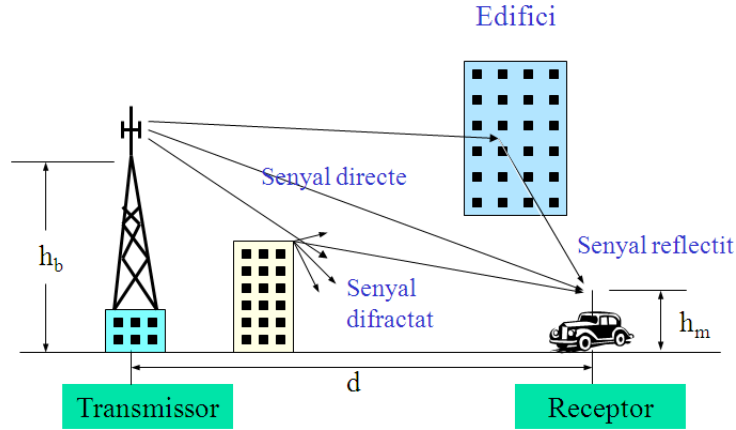


Figura 2.3: Recorregut d'un senyal amb múltiples camins.

nuls o fortes atenuacions i d'altres, amplificacions. Aquest fet fa que es produeixin esvaïments ràpids i de forma aleatòria. A més a més, els objectes o el propi receptor poden estar en moviment i això modifica l'estadística d'aquests esvaïments. Són els anomenats *esvaïments ràpids*. La figura 2.3 il·lustra el recorregut d'un senyal causat per múltiples camins que provoquen instants diferents d'arribada i, per tant, esvaïments.

Si els objectes que causen les reflexions estan pròxims al receptor, els ecos arribaran amb molt poc desfasament respecte el senyal principal, amb poc retard. Són els ecos pròxims i no causen distorsió. És a dir, tot i que modifiquen el senyal principal i sí hi ha un esvaïment, conserva la seva magnitud en totes les components freqüencials. En canvi, si els objectes estan més allunyats, les reflexions arriben amb més desfasament i introdueixen distorsió. Això causa que el canal sigui dispersiu i selectiu en freqüència. L'amplitud d'aquests ecos és aleatòria i està influenciada pels ecos pròxims.

Tot això fa que aquests esvaïments tinguin amplituds i fases diferents en diferents freqüències. El canal és una funció del temps però aquesta funció també és aleatòria en el temps. És a dir, per cada instant de temps es té una funció diferent. La forma de caracteritzar un canal dispersiu és

$$h(\tau, t) = \sum_l A_l(t) \delta(\tau - \tau_l(t)) \quad (2.19)$$

on $A_l(t)$ és l'amplitud complexa de la reflexió l -èssima i $\tau_l(t)$, el seu retard. Cal remarcar que el seu mòdul estadísticament segueix una distribució de Rayleigh i la seva fase segueix una distribució lineal. Així,

$$A_l(t) = |r(t)| e^{j\phi(t)} \quad (2.20)$$

amb $|r(t)| = \sqrt{x(t)^2 + y(t)^2}$ i $\phi(t) = \arctan\left(\frac{y(t)}{x(t)}\right)$, on $x(t)$ i $y(t)$ són dues variables aleatòries gaussianes.

Aquestes reflexions causades per la propagació multicamí fa que el canal respongui com si tingués memòria. La duració d'aquesta memòria o d'aquest pols de resposta s'anomena *delay spread*. Sovint una forma de mesurar aquest paràmetre és avaluar el temps que tarda l'autocorrelació del canal en fer-se menor de 0.5. Aquest paràmetre té un ordre de microsegons.

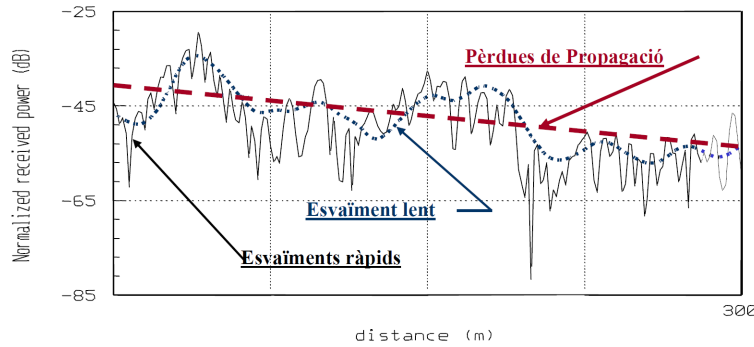


Figura 2.4: Nivell de potència rebut en funció de la distància seguint el model anterior

Finalment, també es produeixen els anomenats *esvaïments lents*. Aquests són causats per la mobilitat del receptor i es produeixen quan, per exemple, travessa per un obstacle de grans dimensions. Durant el temps que travessa aquest obstacle es produeix un decreixement de la potència rebuda. Aquesta variació de potència és molt més lenta que el cas anterior i la probabilitat que es produeixi segueix una estadística lognormal (normal o gaussiana si s'expressa en decibels). La desviació típica d'aquesta variable es correspon amb el *delay spread*. Aquest fet provoca que apareguin zones d'ombra on l'atenuació es veu augmentada. Es coneix amb el nom de *shadowing* i s'utilitza per establir els límits de cobertura per garantir un llindar mínim de sensibilitat.

Així doncs, la funció habitual d'un canal de radiocomunicació comprèn les 3 components anteriors i pren la forma

$$h(\tau, t) = \sqrt{\alpha d^{-\beta} \gamma(t)} \sum_{l=0}^L A_l(t) \delta(\tau - \tau_l(t)) \quad (2.21)$$

on α és una constant, β és la constant de propagació, d és la distància, $\gamma(t)$ és una variable aleatòria lognormal, $A_l(t)$ és l'amplitud de l'eco l -èssim i $\tau_l(t)$, el seu retard. La figura 2.4 il·lustra aquest comportament.

Degut a la naturalesa variant del canal de radiocomunicació, es pot definir un interval temps en el que es pot aproximar la funció del canal sense distorsió. Això equival a dir que durant aquest interval de temps l'amplitud del canal no varia i es manté constant. Aquest temps se l'anomena *temps de coherència*.

Anàlogament, es pot definir un interval freqüencial en el que la resposta freqüencial del canal no tingui distorsió. En aquest marge de freqüències l'amplitud roman plana i totes les freqüències es veuran afectades per igual. Aquest ample de banda s'anomena *ample de banda de coherència*. Això permet introduir el concepte de diversitat freqüencial que es veurà a continuació.

Finalment, si s'estudia l'autocorrelació de l'envoltant, s'observa que a partir de la meitat de la longitud d'ona, l'autocorrelació es fa quasi nul·la. Això equival a dir que, per un instant donat, dos punts espacials separats per una distància superior a la meitat de la longitud d'ona els canals seran independents. Aquest és el principi que explota la diversitat espacial i que es veurà amb profunditat a la secció 2.8.

2.7 Diversitat freqüencial

L'ús de l'ample de coherència permet transmetre informació de manera que n'arribi per diferents freqüències. S'obté el que s'anomena diversitat freqüencial. Si es transmet informació en múltiples portadores freqüencials amb una separació igual o superior a l'ample de banda de coherència, les múltiples freqüències experimentaran esvaïments distints i, com succeïa amb la diversitat temporal, la informació serà fàcilment recuperable. Aquest és el principi que fonamenta la modulació OFDM.

2.7.1 OFDM i OFDMA

La modulació OFDM consisteix en utilitzar múltiples portadores separades correctament i ubicant en cada portadora un conjunt de símbols. En el pla freqüencial, equival a transmetre aquests símbols de forma paral·lela, aconseguint una major velocitat de transmissió. En el pla temporal, el símbol obtingut després d'aplicar la transformada inversa discreta de Fourier conté les característiques suficients com per fer front als esvaïments en freqüència i amb una duració suficient pels esvaïments ràpids. Així doncs, representa una eina molt eficaç i òptima per transmetre a grans velocitats.

Suposi's el cas que es vol assignar un conjunt de portadores OFDM a un usuari. Cada usuari disposarà del seu conjunt de portadores i serà independent de les dels altres usuaris. Es crea una separabilitat entre usuaris en forma de "canals freqüencials" en paral·lel. A més a més, aplicant el concepte de *cross-laying*, es pot utilitzar aquest recurs per donar accés a un usuari. Si no hi ha portadores disponibles, òbviament l'usuari que demana accés no l'obtindrà. És a dir, es crea un gestor d'accés a partir d'un recurs físic. Aquest principi és en el que es basa OFDMA.

Aquesta assignació de portadores als usuaris es pot realitzar de forma preestablerta (cada usuari coneix la seva portadora a priori) o bé, al vol. Cada usuari disposa d'un canal de ràdio independent dels altres usuaris. Això implica que en un instant concret, sempre hi hagi un usuari amb millor canal en una portadora (més amplitud, poques pèrdues) que la resta. Es pot crear un conjunt d'usuaris ordenats de millor a pitjor canal. Així, el millor usuari s'endú l'ús d'aquella portadora. Aplicant aquest concepte a totes les portadores, s'aconsegueix una distribució més òptima de la potència. A més a més, també es pot fixar que el nombre de portadores que pot disposar cada usuari també sigui variable. Els usuaris amb més bons canals, disposaran de més portadores que la resta. La figura 2.5 mostra l'assignació de diferents portadores a diferents usuaris.

No obstant, aquest concepte requereix un gran consum de temps i recursos. A cada instant, l'estació base ha de realitzar aquestes ordenacions però, a més, ha de disposar prèviament la informació del canal de cada usuari. La forma del canal s'obté primerament en recepció aplicant seqüències conegudes per ambdues parts per obtenir així la resposta del canal. És el que es coneix com a CSIR. Un cop l'usuari coneix aquesta resposta, l'envia a l'estació base a través d'un canal paral·lel o *canal de feedback*. L'estació base coneix aquesta resposta i es disposa del CSI en transmissió, el CSIT, i ja pot efectuar l'assignació de recursos i, per tant, dels accessos. Com es pot observar, tot aquest procés requereix temps, potència i una alta capacitat de càlcul, una alta complexitat computacional.

Tot i ser la manera més òptima, actualment és irrealitzable. Per una banda, cada usuari ha d'enviar l'amplitud i fase del canal de cada portadora. Això suposa enviar molta informació a gran velocitat. Habitualment, els terminals mòbils són de poques prestacions i amb

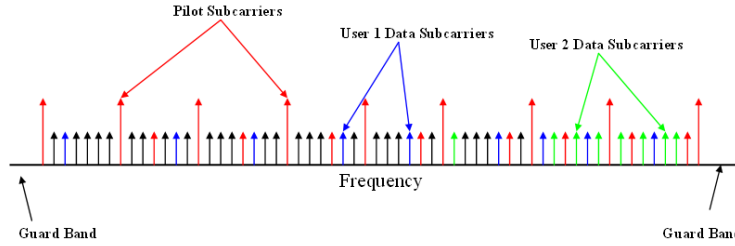


Figura 2.5: Planificació freqüencial d'OFDMA.

una potència molt limitada. Realitzar aquest escenari implicaria que la bateria dels terminals s'esgotés ràpidament. D'altra banda, la complexitat computacional requerida a l'estació base creix exponencialment amb el nombre d'usuaris. Resulta poc realitzable. Així doncs, és necessari buscar una solució subòptima però que sigui més eficaç.

El model de sistema per un usuari k arbitrari resulta

$$\mathbf{y}_k = \mathbf{H}_k \sqrt{\mathbf{P}_k} \mathbf{x}_k + \mathbf{w}_k \quad (2.22)$$

on \mathbf{x}_k és el vector de símbols OFDM per l'usuari k . La matriu $\mathbf{P} = \text{diag}(p_{k,1} \dots p_{k,M})$ és la de potències assignades als símbols de l'usuari k , on $p_{k,m}$ és la potència del seu símbol de la portadora m . Finalment, la matriu de canal \mathbf{H}_k correspon a

$$\mathbf{H}_k = \text{diag}(h_{k,1} \dots h_{k,M}) \quad (2.23)$$

i cada $h_{k,m}$ és el valor aleatori complex del pla freqüencial a la portadora m de l'usuari k .

La modulació OFDM és una candidata prometedora pel futur de les comunicacions sense fils. Ja està implementada en sistemes com el DAB [DAB14], el DVB o el WiFi [IEEE15]. No obstant, també s'utilitza en comunicacions per cable com ADSL [ADS16]. També s'està considerant pel canal de DL per la nova evolució del 3G.

2.8 Diversitat espacial

A diferència de la diversitat freqüencial, la diversitat espacial va començar a utilitzar-se relativament fa pocs anys en sistemes de sonar o radar per tal de millorar la transmissió i recepció d'ones espacialment. Aquesta millora es produeix quan es situen múltiples antenes separades com a mínim la meitat de la longitud d'ona per tal d'obtenir diferents canals incorrelats. Les diferents ones electromagnètiques incideixen amb diferents retards sobre els sensors o antenes i això permet obtenir informació característica.

La inclusió de vàries antenes, tant en transmissió com en recepció, ha obert noves fronteres per explotar aquesta diversitat en espai. A més, si en recepció com en transmissió s'augmenta el nombre d'antenes, s'obté el que s'anomena el canal MIMO amb característiques pròpies que no presenten els altres tipus de canals, com la separació de subcanals o modes paral·lels.

La diversitat espacial, de la mateixa forma que la freqüencial, es basa en l'obtenció o emissió de mostres en diferents punts espaiats i ponderades per diferents amplituds i fases. Aquest fenomen provoca que es creïn sumes constructives i destructives en diferents punts de l'espai, donant lloc a direccions de màxima i mínima potència. Intuïtivament és com si es creés un feix que apuntés cap a una direcció o altra per transmetre o rebre.

A més, si es vol extreure la informació de diferents freqüències n'hi ha prou en crear un filtre temporal FIR a cada sensor per tal que memoritzi les últimes N mostres. Finalment, si posteriorment s'aplica la transformada de Fourier a cada sensor, s'obtindrà l'espectre freqüencial en diferents mostres espacials. Per tant, un conformador en un instant n equival a disposar d'un conformador de banda estreta. Per actuar sobre múltiples freqüències és suficient aplicar una FFT a la sortida. Aquest punt resulta molt òptim per sistemes de multiplexació de banda estreta, com és el cas de OFDMA.

De manera particular, es pot utilitzar la diversitat espacial per “seleccionar” direccions espacials cap on es vol transmetre informació. Aquesta tècnica es coneix com “conformació de feix” o *beamforming*. Es realitza a partir d'un vector que s'anomenarà \mathbf{u} . Així, tenint Q sensors o antenes equispaiades (ULA) el vector d'informació serà $\mathbf{x}_n = [x_1(n) \dots x_Q(n)]^T$. La sortida del conformador serà la mostra $y(n) = \mathbf{u}^H \mathbf{x}_n$.

Cal esmentar que \mathbf{x} és el vector d'ona on queden reflectits els diferents desfasaments que es produeixen degut a la separació d'antenes. Aquests desfasaments són proporcionals a l'angle d'inclinació de l'ona o, essent el mateix, de la direcció. Aquests retards caracteritzen l'ona de forma unívoca ja que determina la posició a l'espai. Així, es pot reescriure \mathbf{x} com

$$\mathbf{x}_n = s(n)\boldsymbol{\vartheta} \quad (2.24)$$

on $s(n)$ és el símbol a transmetre a l'instant n i $\boldsymbol{\vartheta}$ és el vector de retards que determina la direcció. Aquest vector s'anomena “firma espacial” o *steering*.

Una mostra del sensor q del vector d'informació es pot expressar de la forma

$$x_q(n) = s(n)e^{j\omega_0 n}e^{j\theta_q} \quad (2.25)$$

on θ_q equival al retard de fase. Cal tenir present que segons la tipologia del sistema serà d'una forma o altra. No és objectiu d'aquest projecte entrar en detall i anàlisi de tots els tipus de sistemes. El cas que es tractarà en tot el projecte correspon al de sensors equispaiats linealment (ULA) en una dimensió. Per tant, es processa la direcció de l'ona amb l'angle d'inclinació $0 \leq \theta_s \leq 180^\circ$.

Així doncs, pel cas ULA que ocupa aquest projecte,

$$\theta_q = q2\pi f_0 d \frac{\sin(\theta_s)}{c} \quad (2.26)$$

on d és la separació entre antenes, f_0 és la freqüència central i θ_s és la inclinació de la direcció a l'espai. Per tant, per transmetre un símbol $s(n)$ es requereix un vector de retards o desfasaments que caracteritzin una direcció de verticalitat.

Per transmetre Q símbols es requeriran, per tant, Q conformadors ortogonals. Per aquest cas, es pot estendre l'anterior expressió a

$$\mathbf{y}_n = \mathbf{U}^H \mathbf{x}_n \quad (2.27)$$

on $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_Q]$. En aquest cas, el vector d'informació \mathbf{x}_n conté les diferents informacions a l'instant n . D'aquesta manera resulta

$$\mathbf{x}_n = \sum_{q=1}^Q s_q(n)\boldsymbol{\vartheta}_q. \quad (2.28)$$

Noti's que s'escriu $\boldsymbol{\vartheta}_q$ i no $\boldsymbol{\vartheta}_q(n)$ donat que es vol la direcció fixa. Si variés amb el temps, significaria que existeix un moviment. En aquest cas s'hauria d'afegir una component corresponent a l'efecte Doppler. D'altra banda, també es pot escriure la resposta del conformador per Q símbols com

$$y(n) = \text{tr}(\mathbf{U}^H \boldsymbol{\vartheta}) \sum_{q=1}^Q s_q(n). \quad (2.29)$$

on $\boldsymbol{\vartheta} = [\boldsymbol{\vartheta}_1 \dots \boldsymbol{\vartheta}_Q]$ es la matriu que conté els diferents *steerings*.

Cal afegir que es poden realitzar fins a Q conformadors ortogonals. Per $q > Q$, els conformadors no són ortogonals i els senyals es sobreposen espacialment. Així, el cas que suscita interès és per $1 \leq q \leq Q$.

Els pesos del conformador tenen la funció de ponderar la direcció a partir de la introducció de retards i desfasaments per produir sumes constructives i destructives de les ones a l'espai. Per tant, el conformador sol ser estàtic i no en funció de n . Existeixen nombroses tècniques amb objectius diferents per processar aquest fet. És el que es coneix com a *array processing* o processat d'*arrays*.

La conformació no només es pot utilitzar per recepció. En transmissió es pot emprar per dirigir la informació a transmetre cap una direcció en particular. Un esquema típic seria

$$\mathbf{x}_n = \mathbf{U} \sqrt{\mathbf{P}} \mathbf{s}_n \quad (2.30)$$

on \mathbf{U} és la matriu de conformació que conté tots els conformadors \mathbf{u}_q , $\sqrt{\mathbf{P}} = \text{diag}(\sqrt{E_1} \dots \sqrt{E_Q})$ és la matriu de potències diagonal on cada component és l'energia del símbol i $\mathbf{s}_n = [s_1(n) \dots s_Q(n)]^T$ és el vector de símbols unitaris.

La potència d'aquest esquema serà, per tant,

$$P = \mathbb{E} \{ \mathbf{x}_n^H \mathbf{x}_n \} = \mathbb{E} \{ \mathbf{s}^H \sqrt{\mathbf{P}}^T \mathbf{U}^H \mathbf{U} \sqrt{\mathbf{P}} \mathbf{s} \}. \quad (2.31)$$

S'observa clarament que si la potència es gestiona amb la matriu $\sqrt{\mathbf{P}}$, el conformador \mathbf{U} ha de ser unitari. És a dir, el conformador no pot introduir energia al sistema. D'aquesta manera resultarà

$$P = \mathbb{E} \{ \mathbf{s}^H \mathbf{P} \mathbf{s} \} = \text{tr}(\mathbf{P} \mathbb{E} \{ \mathbf{s} \mathbf{s}^H \}). \quad (2.32)$$

Referent al canal $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$, on $N_T = Q$ és el nombre d'antenes en transmissió i N_R és el nombre d'antenes en recepció, pren la forma

$$\mathbf{H} = \begin{bmatrix} h_{11} & \dots & h_{1N_T} \\ \vdots & \ddots & \vdots \\ h_{N_R1} & \dots & h_{N_R N_T} \end{bmatrix} \quad (2.33)$$

on cada h_{ij} correspon al coeficient complex entre l'antena receptora i i la transmissora j . El model de canal és un tema molt ampli de les comunicacions i en constant revisió. Pensar que aquest canal conté les direccions o els vectors de *steering* només correspon a casos molt particulars i que, en cap cas, correspon a la realitat. El model més general és el format per una component fixa i una aleatòria de mitjana nul·la i covariància concreta (canal Rice). Només en aquests casos, la primera component fixa correspondria a entorns on existeix una línia de

visió real entre transmissor i receptor (LOS). La majoria d'escenaris, on no existeix aquesta línia de visió (NLOS), només es disposa de la component aleatòria (canal Rayleigh).

Finalment, pel cas del receptor existeix certa analogia amb el transmissor. Gràcies als teoremes de reciprocitat estudiats a partir de les lleis de Maxwell, el plantejament pel transmissor és anàleg al del receptor, tenint en compte que cal conjugar-ho degut a la visió oposada dels retards.

Així, amb un receptor de N_R antenes, el senyal rebut després del conformador en recepció \mathbf{V} serà

$$\mathbf{y} = \mathbf{V}^H \mathbf{H} \mathbf{U} \sqrt{\mathbf{P}} \mathbf{s} + \mathbf{w} \quad (2.34)$$

Cal tenir molt present que el model algebraic del canal varia en funció del nombre d'antenes i del nombre d'usuaris. Matemàticament no hi ha diferència entre un entorn i l'altre. És a dir, a partir de la matriu \mathbf{H} es pot modelar com si fos un receptor amb N_R antenes i un transmissor amb N_T antenes, N_R usuaris d'una sola antena i un transmissor de N_T antenes, N_R usuaris d'una sola antena i N_T transmissors d'una sola antena, etc. Com es veu, existeixen múltiples combinacions que fan més complex el problema.

2.8.1 Capacitat

De la mateixa manera que prèviament s'ha definit la capacitat per un sistema d'una sola entrada i una sortida (SISO), la capacitat per un sistema de N_T entrades i N_R sortides (canal MIMO punt a punt) resulta una generalització de 2.7. Després de les pertinents manipulacions algebraïques, l'expressió matemàtica més habitual de mostrar la capacitat és

$$C = \log_2 (\det [\mathbf{I}_{N_R} + \mathbf{R}_w^{-1} \mathbf{H} \mathbf{Q} \mathbf{H}^H]) = \log_2 (\det [\mathbf{I}_{N_R} + \mathbf{H}^H \mathbf{R}_w^{-1} \mathbf{H} \mathbf{Q}]) \quad (2.35)$$

on $\mathbf{Q} = \mathbb{E} \{\mathbf{x} \mathbf{x}^H\}$ és la matriu de correlació del vector de transmissió i $\mathbf{R}_w = \mathbb{E} \{\mathbf{w} \mathbf{w}^H\}$ és la matriu de correlació del soroll³. La segona expressió permet mostrar que el canal es pot modelar com si el soroll formés part del canal. Així, definint $\tilde{\mathbf{H}} = \sqrt{\mathbf{R}_w^{-1}} \mathbf{H}$ llavors es tindrà $\mathbf{R}_{\tilde{\mathbf{H}}} = \mathbf{H}^H \mathbf{R}_w^{-1} \mathbf{H}$ de manera que

$$C = \log_2 (\det [\mathbf{I}_{N_R} + \mathbf{R}_{\tilde{\mathbf{H}}} \mathbf{Q}]) = \log_2 (\det [\mathbf{I}_{N_T} + \mathbf{Q} \mathbf{R}_{\tilde{\mathbf{H}}}]) . \quad (2.36)$$

Com es pot apreciar, segons el disseny del transmissor \mathbf{Q} serà una o altra. Per tant, és interessant analitzar quin és el disseny més òptim per maximitzar la capacitat sense incorporar energia addicional.

Suposant que el receptor té un disseny que pot garantir una mínima BER, llavors la solució òptima que maximitza l'expressió anterior es realitza a partir de la descomposició SVD. Pel cas on només hi ha un usuari, partint que el transmissor té tota la informació del canal (és a dir CSIT, coneix $\tilde{\mathbf{H}}$) i suposant soroll blanc, llavors a partir de la descomposició SVD de $\tilde{\mathbf{H}} = \frac{1}{N_0} \mathbf{U}_{\tilde{\mathbf{H}}} \sqrt{\mathbf{D}_{\tilde{\mathbf{H}}}} \mathbf{V}_{\tilde{\mathbf{H}}}^H$ es pot reescriure l'expressió anterior

$$C = \log_2 \left(\det \left[\mathbf{I}_{N_R} + \frac{1}{N_0} \mathbf{U}_{\tilde{\mathbf{H}}} \mathbf{D}_{\tilde{\mathbf{H}}} \mathbf{U}_{\tilde{\mathbf{H}}}^H \mathbf{Q} \right] \right) . \quad (2.37)$$

Si es posa el conformador tal que sigui $\mathbf{U} = \mathbf{U}_{\tilde{\mathbf{H}}}$, llavors $\mathbf{Q} = \mathbf{U}_{\tilde{\mathbf{H}}} \mathbf{P} \mathbf{U}_{\tilde{\mathbf{H}}}^H$ i resulta

³Cal tenir present que si el soroll és blanc llavors $\mathbb{E} \{\mathbf{w} \mathbf{w}^H\} = N_0 \mathbf{I}$.

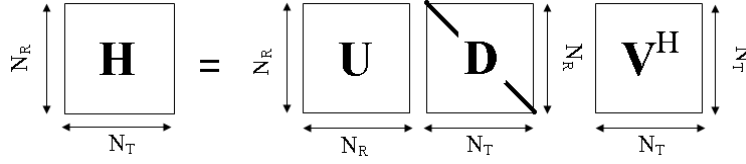


Figura 2.6: Descomposició SVD.

$$\begin{aligned}
C &= \log_2 \left(\det \left[\mathbf{I}_{N_M} + \frac{1}{N_0} \mathbf{U}_{\tilde{\mathbf{H}}} \mathbf{D}_{\tilde{\mathbf{H}}} \mathbf{P} \mathbf{U}_{\tilde{\mathbf{H}}}^H \right] \right) \\
C &= \log_2 \left(\det \left[\mathbf{I}_{N_M} + \frac{1}{N_0} \mathbf{P} \mathbf{D}_{\tilde{\mathbf{H}}} \right] \right) \\
&= \sum_{i=1}^{N_M} \log_2 \left(1 + \frac{1}{N_0} p_i \lambda_{\tilde{\mathbf{H}},i} \right) \\
&\quad s.t. \sum_{i=1}^{N_M} p_i \leq P_T
\end{aligned} \tag{2.38}$$

on p_i és la potència assignada a cada antena en transmissió, $\lambda_{\tilde{\mathbf{H}},i}$ correspon a l'autovalor i de $\tilde{\mathbf{H}}$ i $N_M = \text{rk}(\tilde{\mathbf{H}})$. Aquesta solució quan existeix LOS entre el transmissor i receptor equival al conformador de feix tal que apunta a l'usuari. És a dir, la conformació de feix de la secció anterior no és més que un cas particular de la diagonalització del canal quan hi ha LOS.

Si es vol maximitzar la solució, s'obté que

$$p_i = \left[\mu - \frac{1}{\lambda_{\tilde{\mathbf{H}},i}} \right]^+ = \left[\frac{P_T + \sum_{j=1}^{N_M} \frac{1}{\lambda_{\tilde{\mathbf{H}},j}}}{N_M} - \frac{1}{\lambda_{\tilde{\mathbf{H}},i}} \right]^+ \tag{2.39}$$

on $[x]^+ = \max(0, x)$. Aquesta solució correspon a l'anomenat *water-filling* i la interpretació que té és la d'omplir inversament els autovalors fins un llindar μ . Aquesta solució és la més òptima i, a més, és de baixa complexitat.

No obstant, requereix, com s'ha indicat, el coneixement complet del canal. El transmissor necessita conèixer tots els coeficients de la matriu $\tilde{\mathbf{H}}$ per tal de utilitzar com a conformador els autovectors del canal i assignar la potència a cada antena segons la solució de *water-filling* anterior. Aquests coeficients són estimats pel receptor a partir d'una seqüència d'entrenament i, per tant, són enviats cap el transmissor mitjançant un canal de retorn o un canal de *feedback*. Tot i així, si aquest procés s'ha de realitzar milers de vegades per segon (tenint en compte el temps de coherència del canal), resulta massa complex i no és viable. A més, suposa un consum de potència considerable per part del receptor. Habitualment els receptors acostumen a ser mòbils i amb una potència limitada per la capacitat de la bateria. Inclou, a més, la transmissió dels coeficients del canal suposa un consum de potència addicional que provoca que la bateria s'esgoti de seguida.

Existeixen nombrosos estudis que es dediquen a trobar tècniques i algoritmes per reduir aquesta quantitat de paràmetres a retransmetre o quantitat de *feedback* a canvi de trobar una solució subòptima. Aquest projecte pretén mostrar una solució per l'assignació de potència de baixa complexitat amb un *feedback* reduït.

Aquesta formulació proposada fa referència a comunicacions punt a punt. Només s'inclouen un transmissor i un receptor. Òbviament, aquest no és l'escenari que es contempla, ja que el *broadcasting* requereix múltiples usuaris. Per tant, cal adaptar les expressions anteriors a àmbits on hi hagi múltiples usuaris. El següent capítol es centra en plantejar aquest aspecte a partir del plantejat en els anteriors.

3 Sistemes MIMO Multiusuari

Amb la incorporació de múltiples antenes tant en recepció com en transmissió i el compartiment del medi per part de múltiples usuaris resulta necessari dedicar un capítol a plantejar tota la problemàtica. Els entorns multiusuari suposen una dificultat afegida respecte els entorns punt a punt. A part d'augmentar la complexitat degut al nombre de paràmetres aleatoris, també introdueix el concepte d'interferència multiusuari que queda reflectit amb la SINR.

El tema que ocupa aquest projecte és l'escenari de *broadcasting*. És a dir, l'estació base amb N_T antenes que transmet als usuaris. Aquest tipus d'escenari també se'l coneix com a *Downlink* MIMO Multiusuari.

Matemàticament es pot estendre el canal MIMO punt a punt a un canal MIMO multiusuari. Redefinint el nombre d'antenes en recepció $N_R = \sum_{k=1}^K N_{R,k}$ on $N_{R,k}$ és el nombre d'antenes de l'usuari k , es pot escriure l'equació

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \in \mathbb{C}^{N_R} \quad (3.1)$$

amb els senyals

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_K \end{bmatrix} \in \mathbb{C}^{N_R \times N_T} \quad (3.2)$$

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_K \end{bmatrix} \in \mathbb{C}^{N_T} \quad (3.3)$$

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_K \end{bmatrix} \in \mathbb{C}^{N_R} \quad (3.4)$$

on cada índex fa referència a l'usuari k . De manera particular, $\mathbf{x}_k \in \mathbb{C}^{S_k}$, $\mathbf{H}_k \in \mathbb{C}^{N_{R,k} \times N_T}$, $\mathbf{y}_k \in \mathbb{C}^{N_{R,k}}$ i $\mathbf{w}_k \in \mathbb{C}^{N_{R,k}}$. S_k és el nombre de símbols que corresponen a l'usuari k . Generalment $S_k = 1$ i satisfà $\sum_{k=1}^K S_k = N_T$ i $S_k = N_{R,k}$.

Amb aquest canal MIMO multiusuari es posa de manifest la complexitat a l'hora de diagonalitzar el canal sencer. Es tracta de diagonalitzar cada canal de cada usuari. A part de la gran quantitat de *feedback* que cal retransmetre cap a l'estació base, el càlcul s'incrementa estrepitosament i fa que aquesta opció, tot i ser la més òptima i ser considerada com a cota o *benchmark*, no sigui viable.

El càlcul que es deriva a partir de les generalitzacions anteriors és un càlcul tensorial i tot el problema es pot tractar amb les eines tensorials. Aquest anàlisi queda àmpliament reflectit en els treballs de Martin Haardt, com [HAA05] o [HAA06]. A més a més, cal tenir present que fins ara tot el plantejament correspon a banda estreta i que per múltiples freqüències existeixen múltiples conformadors i múltiples canals, i amb diversitat temporal. És a dir, que en realitat $\tilde{\mathbf{H}}_k \rightarrow \tilde{\mathbf{H}}_k(t, \tau)$. No obstant, per comoditat i simplicitat a l'hora de l'escriptura s'ometran les variables temporals si no són necessàries. Així, encara es pot generalitzar més estenent les equacions anteriors i contemplant les múltiples freqüències. Tot plegat fa que la solució òptima sigui inviable de trobar a dia d'avui.

En termes de capacitat, a diferència del punt anterior on no es contemplava un escenari amb diversos usuaris, queda limitada tal i com succeïa a 3.8. Amb múltiples usuaris cal, però, tenir en compte que es produeixen interferències.

La capacitat del sistema és la que es correspon amb 2.36 però amb el model de canal $\mathbf{R}_{\tilde{\mathbf{H}}}$ i de la matriu \mathbf{Q} a partir de les definicions anteriors. Optimitzar l'expressió de la capacitat amb aquests models, generalment, no sempre és convex. Es fa difícil trobar un punt d'equilibri o expressió tancada. És un problema que segueix obert a dia d'avui.

A part de la diagonalització del canal, existeixen altres estratègies per fer front a altres aspectes del canal MIMO de radiodifusió, com millorar la SINR o explotar la diversitat multiusuari.

Les següents seccions dediquen el seu estudi a mostrar els diferents escenaris de transmissió: MAC i BC. Donat que aquest projecte està dedicat a escenaris de *broadcasting*, aquest s'analitzarà amb més detall que l'altre.

3.1 Accés múltiple

L'accés múltiple (MAC) és aquell en el que varis usuaris volen competir pel medi i enviar la informació a una estació base perquè la rebi. Resulta obvi que és necessari realitzar un ordre, prioritat o ubicar aquestes informacions de tal manera que no es sobreposin i puguin ser processades correctament.

Un dels problemes que se'n deriven és l'obtenció d'una velocitat acceptable de transmissió. A part de la fórmula de Shannon sobre la capacitat 2.7, en un conjunt d'usuaris la capacitat resulta modificada degut a la interferència que es produeix entre ells. La capacitat d'un usuari es veurà afectada per la de tots els altres i així successivament. Així doncs, la capacitat global del sistema resulta de la intersecció de totes les restriccions. Aquesta regió comuna per a tots descriu les velocitats que es poden assolir i que el canal pot absorbir.

Suposi's el cas de dos usuaris que comparteixen el medi. També suposis absència de canal o que el canal pren la forma de $h(t) = \delta(t)$ ¹ (canal AWGN). Considerant el cas SISO ($N_T =$

¹ $\delta(t) * x(t) = x(t), \quad |h(t)|^2 = 1.$

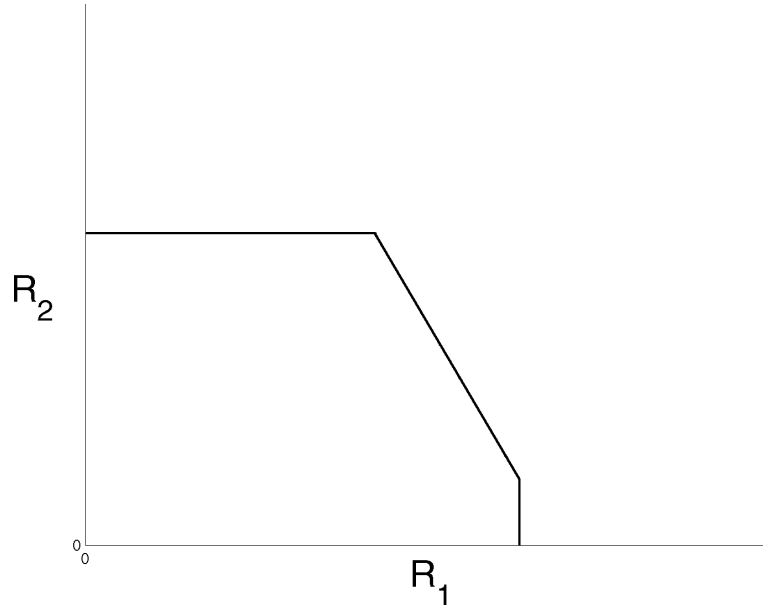


Figura 3.1: Regió de capacitat de MAC.

$N_R = 1$), el senyal rebut a l'estació base pren la forma

$$y(t) = x_1(t) + x_2(t) + w(t) \quad (3.5)$$

on $x_1(t)$ i $x_2(t)$ són els senyals dels usuaris respectius i $w(t)$ el soroll. La capacitat global o *sum capacity* és el màxim cabdal d'informació que es pot aconseguir i correspon a la tercera expressió de les condicions

$$R_1 < \mathcal{I}(x_1; y|x_2) = \log_2 \left(1 + \frac{P_1}{N_0} \right) \quad (3.6)$$

$$R_2 < \mathcal{I}(x_2; y|x_1) = \log_2 \left(1 + \frac{P_2}{N_0} \right) \quad (3.7)$$

$$R_1 + R_2 < \mathcal{I}(x_1, x_2|y) = \mathcal{I}(x_1; y) + \mathcal{I}(x_2; y|x_1) = \log_2 \left(1 + \frac{P_1 + P_2}{N_0} \right). \quad (3.8)$$

Les dues primeres equacions delimiten la capacitat individual de cada usuari. La tercera marca la capacitat conjunta unint les dues potències, el que es coneix com a *sumrate*. Així, es té que

$$C_{sum} = \max \text{sumrate}. \quad (3.9)$$

Unint les condicions de 3.8 es crea una regió de velocitats que es poden aconseguir. És coneix com la regió de capacitat o *capacity region*. Aquesta regió resultant es pot observar a la figura 3.1.

Generalitzant-ho per K usuaris es té

$$\sum_{k \in S} R_k < \log_2 \left(1 + \frac{\sum_{k \in S} P_k}{N_0} \right) \quad \forall S \subset \{1, \dots, K\}. \quad (3.10)$$

3.2 Radiodifusió

Aquest cas es pot considerar l'oposat a l'anterior. L'estació base transmet a diferents usuaris. Aquest escenari resulta diferent perquè la informació que rep un usuari sobre la resta és considerada una interferència i, per tant, no és informació rellevant. Així doncs, l'expressió de la seva capacitat es veu reduïda tant com el nombre d'usuaris. Contemplant un canal AWGN, el senyal rebut per un usuari k arbitrari està compost pel seu senyal i pel de la resta d'usuaris. És a dir,

$$y_k(t) = x_k(t) + \sum_{k' \neq k} x_{k'}(t) + w(t). \quad (3.11)$$

Mentre pel cas anterior de l'accés múltiple per aconseguir la màxima capacitat global era necessari que tots els usuaris transmetessin alhora, per aquest cas de radiodifusió o *broadcasting* s'aconsegueix el màxim quan només es transmet a un usuari i aquest és el que té millor SNR. Aquest projecte centra el seu interès en aquest tipus d'escenari.

Un dels avantatges de l'ús de la diversitat espacial és la de millorar les comunicacions sense, òbviament, afegir més potència al sistema. El més senzill és col·locar múltiples antenes sense realitzar cap processament addicional, donat que es captarà més energia.

El guany per conformació és el que utilitzar un conformador per “apuntar” cap a una direcció en concret. La combinació coherent de les ones electromagnètiques fa que es produeixin màxims i mínims de potència segons la posició de l'espai. Aquest fet fa que la SINR millori perquè pot fer disminuir les interferències situant mínims a la direcció on es situen.

Un dels avantatges és que per produir-se aquesta millora n'hi ha prou que un dels dos extrems de la comunicació (transmissor o receptor) disposi de múltiples antenes, fet que no passava amb el punt anterior on es diagonalitzava el canal.

En canvi, el guany per diversitat consisteix en enviar múltiples rèpliques del mateix senyal que seguiran camins diferents i, en recepció, es disposaran de diferents “versions” del mateix senyal. Aquest punt resulta molt útil quan el canal és molt aleatori, com és el cas dels canals de radiodifusió. L'atenuació que es pot produir en un canal és aleatòria i depèn de múltiples aspectes descrits en el capítol 2. Així, intuïtivament es pot apreciar que en un instant donat serà molt improbable que, donades N_T direccions diferents, es produeixi una atenuació important en totes. El més probable és que alguna d'aquestes no n'hi hagi o sigui poc apreciable respecte les altres. Per tant, quantes més antenes més probable serà trobar un canal amb poca atenuació.

Finalment, el guany per multiplexat espacial és el comentat anteriorment que es produeix quan es diagonalitza el canal. Aquest punt fa que es disposin de $\text{rk}(\tilde{\mathbf{H}})$ canals independents sense interferència amb una potència pròpia, corresponent als autovalors. Aquest guany, però, està sotmès al rang del canal i, per tant, exigeix que hi hagi múltiples antenes tant en recepció com en transmissió i per igual. És a dir, no n'hi ha prou en augmentar el nombre d'antenes a l'extrem de la comunicació. Òbviament resulta molt complex, tant pel càlcul com per termes de costos econòmics.

A continuació es veuran les estratègies de transmissió més habituals que exploten un dels guanys anteriors.

3.2.1 Estratègia òptima

L'estratègia òptima per resoldre aquest problema és coneguda amb el nom de *Dirty Paper Coding*. La solució al problema plantejat anteriorment ve donada per la descodificació successiva als receptors. Però partint del supòsit que els receptors no són cooperatius (un terminal no té per què tenir informació sobre un altre), cada receptor ha de descodificar el seu símbol instantàniament. És a dir, l'estació base ha de precancel·lar les interferències que es produiran. Ha de codificar un símbol a partir dels anteriors, tractant-los com si fossin interferències. Així doncs, el transmissor realitza una codificació successiva i els receptors perceben la comunicació com si fos punt a punt, sense interferències.

El DPC fou publicat per Costa a [COS09]. La idea general és molt senzilla. El transmissor primerament genera el símbol de l'usuari 1. Després, quan genera el símbol de l'usuari 2 té en compte el generat per l'usuari 1 i el pren com si fos interferència. D'aquesta manera, el símbol del segon usuari és el que li pertoca amb el símbol de l'usuari 1 sostret. I així, successivament. El símbols es generen a partir dels anteriors.

D'ençà s'han escrit nombrosos articles que refinen o generalitzen el DPC. Per exemple, [SHM21] elabora un estudi sobre la regió de capacitat del DPC que es correspon amb l'òptima i per què. Caracteritza ambdues regions i demostra que són coincidents.

Definint $\boldsymbol{\pi}$ com el vector d'una permutació d'usuaris, on $\pi(1)$ és l'usuari que es codifica primer, $\pi(2)$ el segon, etc., Π com el conjunt de totes les possibles permutacions de $\boldsymbol{\pi}$, $\mathbf{Q} = [\mathbf{Q}_1 \dots \mathbf{Q}_K]$ i el subespai $\Phi = \left\{ \mathbf{Q} : \text{tr} \left(\sum_{k=1}^K \mathbf{Q}_k \right) \leq P_T \right\}$, la velocitat d'un usuari en una permutació $\boldsymbol{\pi}$ donada és

$$R(\boldsymbol{\pi}, \mathbf{Q}_i) : R_{\pi(i)} = \frac{1}{2} \log_2 \frac{\left| \tilde{\mathbf{H}}_i \left(\sum_{k=i}^K \mathbf{Q}_k \right) \tilde{\mathbf{H}}_i^H + \mathbf{I} \right|}{\left| \tilde{\mathbf{H}}_i \left(\sum_{k=i+1}^K \mathbf{Q}_k \right) \tilde{\mathbf{H}}_i^H + \mathbf{I} \right|}. \quad (3.12)$$

Aquesta estratègia també permet marcar una cota o trobar la capacitat del canal. Així, el *sum capacity* esdevé

$$\begin{aligned} C_{sum} &= \max \sum_{i=1}^K R_i \\ &= \max_{\mathbf{Q}_i} \log_2 \left| \mathbf{I} + \tilde{\mathbf{H}}_1^H \mathbf{Q}_1 \tilde{\mathbf{H}}_1 \right| + \dots + \log_2 \frac{\left| \mathbf{I} + \tilde{\mathbf{H}}_K \left(\sum_{i=1}^K \mathbf{Q}_i \right) \tilde{\mathbf{H}}_K \right|}{\left| \mathbf{I} + \tilde{\mathbf{H}}_K^H \left(\sum_{i=1}^{K-1} \mathbf{Q}_i \right) \tilde{\mathbf{H}}_K \right|} \\ &= \max_{\mathbf{Q}_i} \log_2 \left| \mathbf{I} + \sum_{i=1}^K \tilde{\mathbf{H}}_i^H \mathbf{Q}_i \tilde{\mathbf{H}}_i \right| \\ s.t. \quad &\sum_i \text{tr}(\mathbf{Q}_i) \leq P_T \end{aligned} \quad (3.13)$$

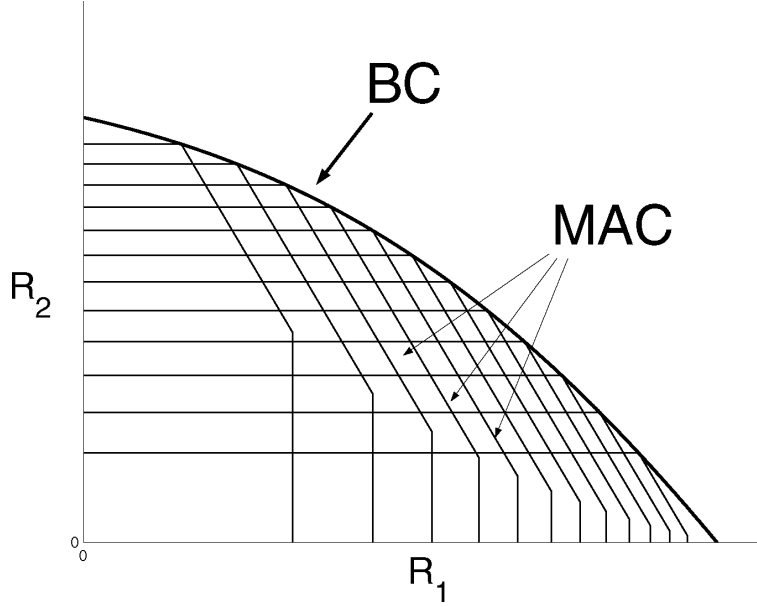


Figura 3.2: Regió de capacitat de BC.

La regió de capacitat, per tant, és la unió convexa de totes les velocitats que es poden aconseguir per totes les permutacions possibles d'usuaris i totes les matrius definides positivament de correlació \mathbf{Q}_k que compleixin la restricció de potència. Llavors es pot escriure que la capacitat resulta

$$C = \text{Co} \left(\bigcup_{\pi, \mathbf{Q}} R(\pi, \mathbf{Q}) : \mathbf{Q} \in \Phi, \pi \in \Pi \right). \quad (3.14)$$

No obstant, resulta molt complex trobar aquesta regió degut a les moltes possibles combinacions de les matrius \mathbf{Q}_i . Aquest fet degenera en problemes no convexos i fa que sigui necessari trobar una expressió equivalent. D'aquesta manera [GOL10] proposa trobar la regió de capacitat a partir de la dualitat amb el canal de *uplink*. Aquest canal correspon al que perceben els usuaris quan transmeten cap a l'estació base (el cas MAC). El lector ja pot entreveure que el canal, per aquest escenari oposat, és

$$\mathbf{H}_k^{BC} = (\mathbf{H}_k^{MAC})^H. \quad (3.15)$$

Així, la regió de capacitat del canal BC serà la unió de les regions de capacitat dels canal MAC. Escrit d'altra forma és

$$C^{BC}(P_T, \mathbf{H}_1, \dots, \mathbf{H}_K) = \bigcup_{\mathbf{G}_k: \sum_{k=1}^K \text{tr}(\mathbf{G}_k) \leq P_T} C^{MAC}(\mathbf{G}_1, \dots, \mathbf{G}_K, \mathbf{H}_1^H, \dots, \mathbf{H}_K^H) \quad (3.16)$$

on \mathbf{G}_i és la matriu de potències de l'usuari i per l'escenari MAC. La figura 3.2 il·lustra gràficament aquesta unió. La regió de capacitat resultant és la corba envoltant.

L'assignació òptima de potència ve donada per iteracions i recerques exhaustives per maximitzar l'expressió de capacitat, que posen de manifest una alta complexitat, tal com s'aprecia a [JUN11] i [GOL12].

3.2.2 Implementacions òptimes

La solució que presenta el DPC resulta ser l'òptima que satisfà la maximitzava de la capacitat. No obstant, la seva implementació no resulta trivial. L'altra complexitat que presenta fa que a dia d'avui sigui irrealitzable. La recerca actual busca noves vies per implementar el DPC d'una forma menys complexa. Tot i això, continua mostrant alta complexitat en els càlculs.

La primera implementació que es mostra en aquest projecte és la que presenta [JUN11]. S'ha escollit perquè resulta atractiva perquè presenta un model de sistema on el càlcul matricial inclou la diversitat freqüencial i espacial. És a dir, la matriu de correlació que s'ha de maximitzar \mathbf{Q} inclou ambdues diversitats.

La manera que ofereix per implementar-ho és la de resoldre les matrius òptimes \mathbf{Q}_k de cada usuari per parelles. És a dir, en cada iteració optimitza un parell de matrius. Així la complexitat es redueix i les següents iteracions es basen en els resultats anteriors. Tot i això, trobar les matrius \mathbf{Q}_k que maximitzen el *sumrate* continua essent una tasca complexa.

Una altra implementació interessant que s'ha utilitzat com a *benchmark* en aquest projecte és la que proposa [KOB08]. S'ha escollit aquesta perquè resol el problema DPC però hi incorpora pesos o prioritats per garantir una QoS. Per fer-ho pren la dualitat entre MAC i BC i resol el problema en MAC. Després aplica la dualitat i obté la solució dual per BC. No obstant, tot i que no és rellevant pel títol d'aquesta secció, la solució que aporta no contempla la diversitat freqüencial.

El que proposa és maximitzar l'expressió

$$\max_{\mathbf{p}} \sum_{k=1}^K \phi_{\pi_k} \log_2 \frac{\left| \mathbf{I} + \sum_{j=1}^k \mathbf{h}_{\pi_j} \mathbf{h}_{\pi_j}^H p_{\pi_j} \right|}{\left| \mathbf{I} + \sum_{j=1}^{k-1} \mathbf{h}_{\pi_j} \mathbf{h}_{\pi_j}^H p_{\pi_j} \right|} \quad (3.17)$$

que es correspon amb l'expressió del DPC per MAC incorporant-hi les prioritats. Aquestes prioritats ϕ_{π_i} segueixen un ordre decreixent tal que

$$\phi_{\pi_1} > \dots > \phi_{\pi_K} \quad (3.18)$$

i els usuaris es descodifiquen amb K primer i 1, l'últim. Així doncs, el problema es pot reescriure de la manera

$$\begin{aligned} \max_{\mathbf{p}} f(\mathbf{p}) &= \max_{\mathbf{p}} \sum_{k=1}^K \Lambda_k \log_2 \left| \mathbf{I} + \sum_{j=1}^k \mathbf{h}_j \mathbf{h}_j^H p_j \right| \\ s.t. \quad &\sum_{k=1}^K p_k = P \\ &\mathbf{p} \geq \mathbf{0} \end{aligned} \quad (3.19)$$

on P és el llindar de potència i $\Lambda_k = \phi_k - \phi_{k+1}$, definint $\phi_{K+1} = 0$. Per optimitzar l'equació anterior tenint present les restriccions s'aplica el mètode dels multiplicadors de Lagrange. Aquests permeten relaxar el problema i trobar una solució òptima.

Per tant, el lagrangiana del problema anterior resulta

$$L(\mathbf{p}, \lambda, \boldsymbol{\mu}) = f(\mathbf{p}) + \sum_{k=1}^K \mu_k p_k - \lambda \left(\sum_{k=1}^K p_k - P \right) \quad (3.20)$$

on λ i $\boldsymbol{\mu}$ són les variables duals. Per optimitzar el nou problema, només cal diferenciar el lagrangiana respecte p_k . Així doncs, les condicions KKT que se'n deriven són

$$\begin{aligned} \frac{\delta f(\mathbf{p})}{\delta p_k} + \mu_k - \lambda &= 0, \quad 1 \leq k \leq K \\ p_k &\geq 0, \quad 1 \leq k \leq K \\ P - \sum_{k=1}^K p_k &\geq 0 \\ \mu_k &\geq 0, \quad 1 \leq k \leq K \\ \lambda &\geq 0. \end{aligned} \quad (3.21)$$

En aquest punt, cal tenir present la descodificació successiva que introdueix el DPC. Per tant, es pot definir la matriu de covariància de la interferència i el soroll per un usuari k

$$\boldsymbol{\Sigma}_{k,j}(\mathbf{p}) = \mathbf{I} + \sum_{i=1, i \neq k}^j \mathbf{h}_i \mathbf{h}_i^H p_i \quad (3.22)$$

tenint en compte que $j \geq k$ i que els usuaris de K fins $j+1$ ja han estat descodificats. Prenent el problema i a partir de la definició anterior, es pot escriure

$$\sum_{j=k}^K \Lambda_j \frac{\mathbf{h}_k^H \boldsymbol{\Sigma}_{k,j}^{-1}(\mathbf{p}) \mathbf{h}_k}{1 + p_k \mathbf{h}_k^H \boldsymbol{\Sigma}_{k,j}^{-1}(\mathbf{p}) \mathbf{h}_k} = \lambda - \mu_k, \quad 1 \leq k \leq K. \quad (3.23)$$

Resoldre aquest problema no és trivial i no es pot trobar una funció implícita que trobi els valors de p_k a la vegada. Per fer-ho, llavors, es planteja un algorisme iteratiu. Definint $\mathbf{p}^0 = \mathbf{0}$ com el vector de potència a la iteració 0, llavors en una iteració i arbitrària es calculen els coeficients

$$\alpha_{k,j}^i = \mathbf{h}_k^H \boldsymbol{\Sigma}_{k,j}^{-1}(\mathbf{p}^{i-1}) \mathbf{h}_k \quad (3.24)$$

per cada $k = 1, \dots, K$ i $j \geq k$. La solució al problema es pot obtenir aplicant *water-filling* a

$$\tilde{\mathbf{p}}^i = \arg \max_{\mathbf{p}: \sum_k p_k = P} \sum_{j=1}^K \Lambda_j \sum_{k=1}^j \log_2 (1 + \alpha_{k,j}^i p_k). \quad (3.25)$$

La solució a aquest pas es pot trobar a l'apèndix de [KOB08] i s'ometrà per no estendre's. Finalment, s'actualitza el vector de potències a partir de l'expressió

$$\mathbf{p}^i = \frac{1}{K} \tilde{\mathbf{p}}^i + \left(1 - \frac{1}{K}\right) \mathbf{p}^{i-1}. \quad (3.26)$$

D'aquesta manera s'obté un vector de potències òptim que maximitza el *sumrate* amb prioritats. Cal tenir present que aquesta implementació no es centra en el conformador, donat que només es dedica a maximitzar l'expressió del *sumrate* a partir de la matriu de covariàncies $\mathbf{Q}^{MAC} = \mathbb{E}\{\mathbf{p}\mathbf{p}^H\}$. No obstant, cal tenir present que la solució que es troba correspon a l'escenari MAC. Cal trobar el vector de potències en el pla de BC. A partir de la dualitat que presenten [GOL22], es pot trobar la matriu de covariàncies de BC a partir de \mathbf{Q}^{MAC} . Concretament, per simplificar la notació es defineix

$$\begin{aligned}\mathbf{A}_k &= \mathbf{I} + \mathbf{H}_k \left(\sum_{j=1}^{k-1} \mathbf{Q}_j^{BC} \right) \mathbf{H}_k^H \\ \mathbf{B}_k &= \mathbf{I} + \sum_{j=k+1}^K \mathbf{H}_j^H \mathbf{Q}_j^{MAC} \mathbf{H}_j.\end{aligned}\tag{3.27}$$

Lavors es pot escriure que la matriu de covariància és

$$\mathbf{Q}_k^{BC} = \mathbf{B}_k^{-1/2} \mathbf{F}_k \mathbf{G}_k^H \mathbf{A}_k^{1/2} \mathbf{Q}_k^{MAC} \mathbf{A}_k^{1/2} \mathbf{G}_k \mathbf{F}_k^H \mathbf{B}_k^{-1/2}\tag{3.28}$$

on les matrius \mathbf{F}_k i \mathbf{G}_k corresponen a la descomposició SVD del canal efectiu $\mathbf{B}_k^{-1/2} \mathbf{H}_k^H \mathbf{A}_k^{1/2} = \mathbf{F}_k \mathbf{\Lambda}_k \mathbf{G}_k^H$.

Així doncs, un cop es té la matriu de covariàncies en BC, les components del vector de potències són les arrels quadrades de la diagonal de la matriu de covariància. És a dir,

$$p_k^{BC} = \sqrt{Q_{kk}^{BC}}.\tag{3.29}$$

Aquest algoritme resol el plantejament del DPC de manera que s'obté un vector de potències on cada componen fa referència a un usuari. No obstant, cal recordar que no fa ús de la diversitat freqüencial. Noti's que \mathbf{Q}_k^{BC} només depèn de les anteriors $k - 1$ matrius. Per tant, s'han de calcular en l'ordre que s'ha establert.

3.2.3 Implementacions subòptimes

Com s'ha vist anteriorment, la solució òptima requereix una alta complexitat, difícil d'implementar amb la tecnologia actual. Els esforços de la comunitat científica es centren en gran part a trobar solucions subòptimes però que ofereixin un resultat força pròxim a l'objectiu òptim i, alhora, de menys complexitat.

A continuació es descriuen les implementacions més habituals.

3.2.3.1 Descomposició RQ

Una manera d'implementar la cancel·lació d'usuaris és aplicant una descomposició triangular al canal. Concretament es proposa utilitzar la descomposició RQ del canal². D'aquesta manera queda

$$\mathbf{H} = \mathbf{R}\mathbf{Q}_H^H\tag{3.30}$$

²Noti's que la matriu \mathbf{Q}_H no té res a veure amb $\mathbf{Q} = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\}$.

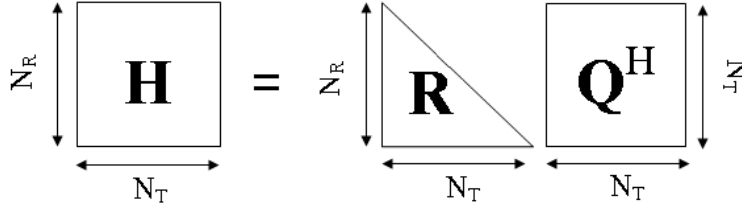


Figura 3.3: Descomposició RQ.

on $\mathbf{R} = \begin{bmatrix} R_{11} & 0 & 0 & \cdots & 0 \\ R_{21} & R_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ R_{N_R 1} & R_{N_R 2} & R_{N_R 3} & \cdots & R_{N_R N_T} \end{bmatrix} \in \mathbb{C}^{N_R \times N_T}$ és una matriu triangular inferior, i $\mathbf{Q}_H \in \mathbb{C}^{N_T \times N_T}$ és una matriu unitària, de forma que $\mathbf{Q}_H^H \mathbf{Q}_H = \mathbf{I}$. Així, si es posa el conformador de tal forma que $\mathbf{U} = \mathbf{Q}_H$ llavors la comunicació serà

$$\mathbf{r} = \mathbf{R} \mathbf{Q}_H^H \mathbf{Q}_H \sqrt{\mathbf{P}} \mathbf{s} + \mathbf{w} = \mathbf{R} \sqrt{\mathbf{P}} \mathbf{s} + \mathbf{w} \quad (3.31)$$

i així és el propi canal qui realitza la codificació successiva, ja que

$$\begin{aligned} r_1 &= R_{11} \sqrt{p_1} s_1 + w_1 \\ r_2 &= R_{21} \sqrt{p_1} s_1 + R_{22} \sqrt{p_2} s_2 + w_2 \\ &\dots \\ r_{N_R} &= R_{N_R 1} \sqrt{p_1} s_1 + R_{N_R 2} \sqrt{p_2} s_2 + \dots + R_{N_R N_T} \sqrt{p_{N_T}} s_{N_T} + w_{N_R}. \end{aligned} \quad (3.32)$$

Falta, però, la selecció d'usuaris. Sabent que com a màxim es poden servir N_T usuaris, si $N_T < K$ llavors caldrà realitzar una selecció prèvia dels usuaris. Per tant, a més de requerir un conformador particular també és necessari efectuar una prèvia assignació d'accés. Per tant, cal disposar de tota la informació del canal de cada usuari, encara que posteriorment siguin descartats. La complexitat, com es veu, que creix escandalosament amb el nombre d'usuaris.

Com ja es pot apreciar, aquesta implementació resulta subòptima perquè no realitza ni una selecció òptima d'usuaris ni ordenació adequada. A diferència dels esquemes plantejats anteriorment on es realitzava una maximització, la descomposició RQ no ofereix un resultat màxim per la capacitat.

Una interpretació intuïtiva de la descomposició RQ és la de trobar una transformació que canviï el canal en una matriu triangular i aplicar la precancel·lació successiva. D'aquesta manera el canal resultant és triangular i aplica ja per sí sol la descodificació successiva, tal i com mostra la figura 3.4.

3.2.3.2 Forçador de zeros

Una altra estratègia que utilitza la cancel·lació d'interferències és el forçador de zeros. Consisteix en forçar màxims a les direccions on es troben els usuaris als que se'ls dona accés. Per tant, no es produeix una cancel·lació successiva i no es correspon amb el DPC. Aquesta estratègia, tot i ser subòptima, permet cancel·lar les interferències de forma que cada usuari

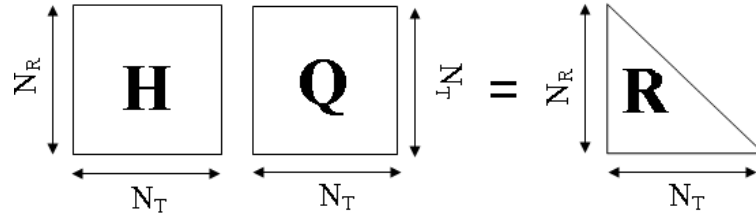


Figura 3.4: Aplicació RQ.

disposi d'un feix per ell de manera que percebi el seu símbol amb la màxima potència i amb mínima interferència els dels altres.

Aquest fet pot millorar la SINR d'un usuari, ja que cancel·la les interferències del denominador i incrementa el resultat total. No obstant, té el contraefecte d'augmentar el soroll.

Així doncs, el conformador es realitza a partir del canal MIMO de manera que

$$\mathbf{U} = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1} \quad (3.33)$$

tenint en compte que \mathbf{H} és el canal dels usuaris seleccionats prèviament. En sí es tracta de la pseudo-inversa de \mathbf{H} per tal de cancel·lar-lo. Substituint quedarà

$$\mathbf{r} = \mathbf{H}\mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1} \sqrt{\mathbf{P}}\mathbf{s} + \mathbf{w} = \sqrt{\mathbf{P}}\mathbf{s} + \mathbf{w}. \quad (3.34)$$

Realment ha cancel·lat l'efecte del canal, com si d'un equalitzador es tractés. No obstant, no explota la diversitat multiusuari. En un moment donat, pot ser que la potència del canal, expressada per $\text{tr}(\mathbf{H}\mathbf{H}^H)$, sigui important. El forçador de zeros sempre cancel·larà aquest efecte, inclòs quan sigui profitós. Per tant, resulta una estratègia subòptima. La següent estratègia reflecteix com explotar aquest punt a favor. Es pot estudiar a [HAA13].

Aquesta estratègia, com s'ha comentat, pot millorar la SINR degut a l'eliminació de les interferències. No obstant, no fixa una política de selecció d'usuaris i, a més, requereix el coneixement total del canal (CSIT) de tots els usuaris.

3.2.3.3 Conformació oportunista

La política oportunista (MOB) va ser introduïda per [VTL17] i ha sorgit amb força els últims anys. Intuïtivament equival a seleccionar aquells usuaris que disposin d'un millor canal, és a dir, de forma oportunista. A més, s'ha demostrat que resulta ser la solució òptima al problema. Seleccionar aquells usuaris que disposin d'un millor canal, farà que el *sumrate* es vegi incrementat. Tot i així, té l'inconvenient que depèn estretament del nombre d'usuaris per millorar la capacitat. Resulta evident que amb molts usuaris serà molt més probable que n'hi hagi un grup que disposin d'un bon canal; en canvi, amb pocs usuaris, possiblement es trobaran més pocs punts de baixa atenuació.

Aquesta filosofia es pot aplicar a molts paràmetres de selecció. En freqüència equivaldria a seleccionar aquelles portadores amb bona amplitud dels millors usuaris. Dit d'altra manera, per cada portadora, seleccionar l'usuari amb millor amplitud. Anàlogament, en temps correspondria a seleccionar, en un instant donat, l'usuari que disposés d'un millor canal.

Però, què vol dir disposar d'un millor canal? Segons el punt de vista aquest concepte serà diferent. Per exemple, en termes de capacitat, el millor canal és el que aporta més velocitat. I

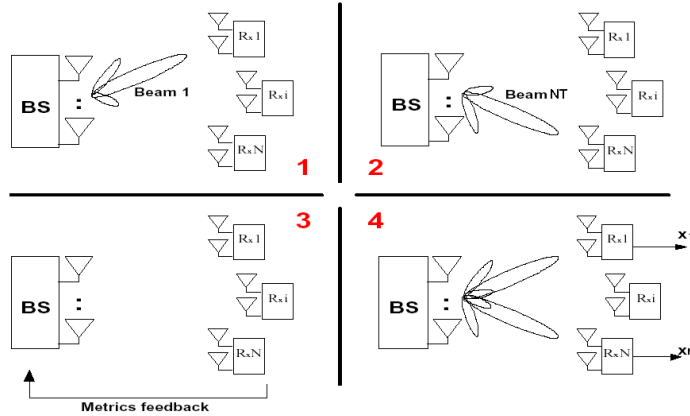


Figura 3.5: Estratègia MOB.

encara més, es podria veure que el millor canal és aquell que incrementa el *sumrate*. En canvi, en termes de BER, el millor canal és aquell que ofereix més protecció davant la taxa d'error. És a dir, un usuari pot tenir un bon canal perquè n'augmenta la velocitat però que sigui un mal canal perquè fa incrementar la BER.

En aquest projecte s'estudia des del punt de vista de maximitzar el *sumrate* del sistema. Però, a més, també cal apreciar quin paràmetre es pot extreure per avaluar quant bo és un canal. En sistemes punt a punt d'una sola antena (SISO), aquest paràmetre resulta obvi donat que senzillament serà $|h(t)|^2$. En sistemes MISO o SIMO (típicament usuaris d'una sola antena i una estació base de múltiples antenes)³ també resulta fàcil apreciar que el *sumrate* és proporcional al conjunt d'usuaris amb millor norma de canal, $\|\mathbf{h}\|^2$. No obstant, en sistemes MIMO aquest punt es complica donat que no existeix una manipulació matemàtica que ofereixi la qualitat d'un canal. Aquest tema ha estat àmpliament estudiat i suposa un punt important per reduir el *feedback*.

Així doncs, en escenaris MISO, la selecció d'usuaris s'efectua a partir de quant bo és el canal. Aplicat a la diversitat espacial es pot dissenyar a partir d'un conformador amb múltiples feixos aleatoris i seleccionant els usuaris a partir de la qualitat de senyal rebuda. Així doncs, els passos a seguir són:

1. L'estació base conforma múltiples *beams* ortonormals aleatoris.
2. Els va activant un a un i transmet un senyal conegut pels receptors.
3. Un cop ha acabat, tots els receptors envien a l'estació base el valor de la SINR que han percebut per cada *beam*.
4. L'estació base activa tots els *beams* i transmet els símbols dels usuaris amb millor canal pel millor *beam*.

Aquests passos queden representats a la figura 3.5.

³Cal recordar que aquests sistemes també es poden modelar com a un canal MIMO. No obstant, el canal MIMO farà referència als escenaris amb múltiples antenes tant en recepció com en transmissió.

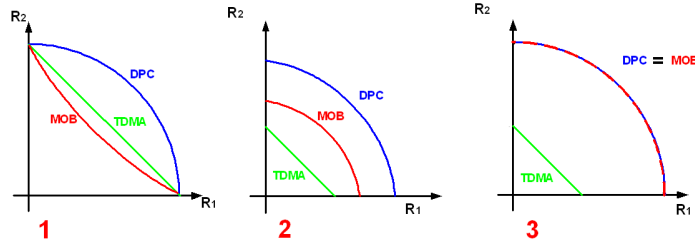


Figura 3.6: Comparació de les estratègies TDMA, MOB i DPC per $K = 2$, $K = 100$ i $K = \infty$ amb $N_T = 2$.

Aquesta estratègia ofereix nombrosos avantatges molt atractius. Primerament permet implementar un disseny molt senzill, de baixa complexitat. L'algoritme anterior només presenta una complexitat de $\mathcal{O}(N_T)$ a l'estació base.

Seguidament, no fa ús del CSI total. És a dir, no li cal conèixer tota la matriu \mathbf{H} ja que només amb la SINR com a paràmetre de *feedback* n'hi ha prou. Òbviament es pot refinar i fer que n'augmentin els paràmetres segons l'estratègia, tal i com es veurà en el següent capítol. Tot i així, continua essent poca quantitat de *feedback*.

En un altre punt hi ha el resultat obtingut. Aquesta estratègia explota la diversitat d'usuari. En altres paraules, tal i com s'ha explicat anteriorment, quants més usuaris hi hagi en el sistema, més probabilitat que hi hagi un bon canal o inclòs que algú tingui visió directa. Per tant, la diversitat d'usuaris fa millorar els resultats. No obstant, amb pocs usuaris ja s'aprecia una millora respecte altres tècniques clàssiques.

Així doncs, per un gran nombre d'usuaris, el *sumrate* s'aproxima a la capacitat del DPC. Per tant, asimptòticament es té que

$$\lim_{K \rightarrow \infty} C_{\text{sum}} = C^{\text{DPC}} = N_T \log_2 \log_2(N_T \cdot K). \quad (3.35)$$

La figura 3.6 compara l'estratègia MOB amb el DPC i el TDMA. L'estratègia TDMA correspon a donar accés als usuaris en instants diferents. La primera representació és pel cas $K = 2$ i s'aprecia com MOB està per sota de TDMA. La representació 2 és per $K = 100$ i s'observa com la regió de capacitat augmenta degut al guany per diversitat d'usuari. Finalment, la representació 3 és pel cas $K \rightarrow \infty$. En aquest cas, les capacitats de DPC i de MOB coincideixen i demostren com és el comportament asimptòtic.

Finalment, com a altre avantatge, des d'un punt de vista de *cross-layering* hi ha l'accés. Aquest mecanisme ja porta una selecció d'usuaris implícitament. Dóna o denega accés segons els paràmetres que han retransmès els usuaris. Així doncs, a part de gestionar els recursos físics en forma de conformador també fa que es produeixi un accés. Per tant, la conformació oportunista actua dins la capa PHY-MAC.

Un altre punt a tenir en compte és el nombre de feixos o *beams*. Com s'ha comentat, només es poden conformar fins a N_T feixos ortogonals entre sí. No obstant, no sempre pot ser la solució més òptima utilitzar N_T *beams*. Un alt nombre de *beams* incrementa la interferència i, per tant, fa disminuir la SINR.

Existeixen nombroses solucions per trobar el nombre òptim d'usuaris a servir o el nombre òptims de *beams*. Es pot realitzar a través d'optimitzacions instantànies o bé per aproximacions estadístiques de la probabilitat d'assolir una SINR lliandar. No obstant, el pròxim capítol oferirà

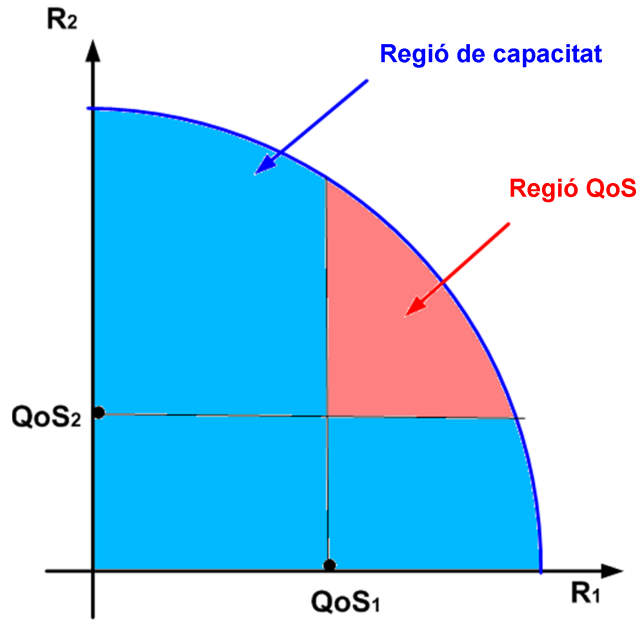


Figura 3.7: Regió QoS.

una altra manera de trobar un valor òptim.

Posar aquest paràmetre com a variable incremental, com és evident, la complexitat del sistema. Donat que els usuaris no coneixen a priori aquest paràmetre, no poden calcular les permutacions que es poden efectuar amb els *beams* que s'han processat. Per tant, és una tasca que ha de realitzar l'estació base i, per tant, no pot distribuir complexitat entre els usuaris.

3.2.4 Regió QoS

Un dels aspectes més rellevants que es desprenen de la gestió d'accés mitjançant el *cross-laying* és l'assignació de prioritats sobre els usuaris. Situant usuaris més prioritaris que d'altres farà que obtinguin una velocitat major. Resulta evident que la descàrrega de veu no requereix la mateixa velocitat que la de vídeo. Així doncs, un usuari que requereixi visualitzar vídeo en temps real demandarà més velocitat que l'usuari que només visualitza webs. És el que s'anomena tràfic heterogeni. Així doncs, hi ha una demanda mínima de qualitat de servei (QoS).

Aquestes restriccions limiten encara més la regió de capacitat. La figura 3.7 mostra com la regió de QoS és la intersecció de les regions QoS de cada usuari. La incorporació de QoS, no obstant, fa que la justícia variï i puguin sorgir injustícies i monopolis. Per exemple, usuaris amb una alta QoS pot ocasionar que demandin tota la velocitat del sistema. Per aquesta raó, el planificador ha de tenir present aquest punt.

El següent capítol presenta una tècnica per maximitzar el *sumrate* però, alhora, guardar unes prioritats per garantir una QoS.



4 Assignació de potència per OFDMA-multiantena

LA creixent demanda dels últims anys de recursos i usuaris ha portat en sí mateix la recerca de noves tècniques per fer-la front. Amb la incorporació de la diversitat espacial en forma de múltiples antenes, els canals MIMO cada cop són més presents a les comunicacions. Així mateix, la gestió d'accés cada cop cobra més rellevància en un àmbit on el número d'usuaris va creixent i les restriccions a l'operador són més estrictes. El *cross-laying* dona resposta aquesta demanda i la recerca que es presenta en aquest capítol vol donar solució a aquest problema.

Actualment, els escenaris MIMO amb múltiples usuaris (MU-MIMO) són una via atractiva per incrementar la velocitat del sistema i es presenta com a un tema cada cop més a l'alça segons es pot llegir a [UTS23], [HAA24] o [GOL25]. Els resultats que ofereixen [GOL25] i [WYU26] ofereixen formes duals per resoldre els problemes. És a dir, plantegen un problema dual amb altres variables però de menys dificultat de resoldre.

Paral·lelament, la multiplexació espacial està essent un recurs molt socorregut i cada cop es té més present per combinar-lo amb sistemes de diversitat freqüencial, tal i com mostren resultats a [JUN11], [OTT27], [WAN28] o [KOU07].

Així doncs, s'aprecia que el tema que a continuació es tracta és un tema actual i sobre el que la comunitat científica hi treballa. El que el lector hi trobarà serà una solució de baixa complexitat que contempla un escenari MIMO-DL en OFDMA. Per fer-ho, a diferència de [JUN11], es proposa una descomposició dual del problema.

4.1 Escenari

L'escenari que es contempla és un canal de *broadcasting* per OFDMA degut al creixent ús per part de les tecnologies recents d'aquesta tècnica. L'objectiu a perseguir és el d'optimitzar el *sumrate* de l'escenari de forma que requereixi poca complexitat. A més, també es contempla l'opció de prioritats o pesos en els usuaris. És a dir, hi hagi usuaris amb més prioritat que d'altres. Per tal d'explotar la diversitat temporal i reduir la complexitat, s'optimitzarà el *sumrate* ergodic¹ en comptes del valor instantani.

El model de senyal considerat és d'una sola cèl·lula OFDMA amb una estació base (s'ignoren els efectes intercel·lulars), K usuaris i M portadores. Cada usuari només té una antena i rep N_T senyals diferents, que poden venir de diferents antenes o altres direccions. Només

¹Equival a optimitzar el valor mig del *sumrate*.

un d'aquests senyals està destinat a un sol usuari (es crea una relació entre usuari i *beam* unívoca). El senyal que rep un usuari k arbitrari a la portadora m -èssima és

$$y_{k,m} = \mathbf{h}_{k,m}^H \sum_{q=1}^{N_T} \mathbf{b}_{q,m} \sqrt{p_{k,q,m}} s_{k,q,m} + \mathbf{w}_{k,m} \quad (4.1)$$

on $\mathbf{h}_{k,m} \in \mathbb{C}^{N_T}$ és el vector de l'usuari k a la freqüència m , $\mathbf{b}_{q,m} \in \mathbb{C}^{N_T}$ és el conformador de feix unitari q -èssim a la freqüència m , $p_{k,q,m}$ és la potència de l'usuari k assignat al *beam* q de la freqüència m —l'índex $k_{q,m}$ denota l'usuari assignat al *beam* q de la freqüència m —, $s_{k,q,m}$ la seva informació i $\mathbf{w}_{k,m}$ el vector de soroll additiu gaussià. Cal notar que la matriu de conformació $\mathbf{B}_m = [\mathbf{b}_{1,m} \dots \mathbf{b}_{N_T,m}]$ és unitària ($\mathbf{B}_m^H \mathbf{B}_m = \mathbf{I}_{N_T}$) i $|s_{k,m}|^2 = 1$. Per comoditat s'obviaran els índex temporals en tot el capítol a excepció que es requereixi.

Com s'aprecia, que sigui un entorn multiantena, multiusuari i multiportadora incrementa el número de paràmetres i graus de llibertat de la capa PHY.

Considerant la matriu de potències $\mathbf{P} \in \mathbb{R}^{+,K \times M} \cup \{0\}$ on cada $p_{k,m}$ correspon a la potència assignada a l'usuari k -èssim a la portadora m -èssima —si $p_{k,m} = 0$ l'usuari no té accés a aquesta portadora—, l'objectiu plantejat amb les restriccions de potència es pot formular de la forma

$$\begin{aligned} \max_{\mathbf{P}} R = \max_{\mathbf{P}} \mathbb{E} \left\{ \sum_{m=1}^M \sum_{k=1}^K \log_2 (1 + \gamma_{k,m}(\mathbf{P})) \right\} \\ \text{s.t. } \mathbb{E} \left\{ \sum_{m=1}^M \sum_{k=1}^K p_{k,m} \right\} \leq \bar{P} \\ \mathbb{E} \left\{ \sum_{m=1}^M \log_2 (1 + \gamma_{k,m}(\mathbf{P})) \right\} \geq \phi_k R, \quad k = 1, \dots, K \end{aligned} \quad (4.2)$$

on \bar{P} és la potència mitja disponible pel sistema, ϕ_k és la prioritat de l'usuari k i $\gamma_{k,m}(\mathbf{P})$ és la SINR que percep l'usuari k -èssim a la freqüència m -èssima. Aquesta última es defineix

$$\gamma_{k,m}(\mathbf{P}) = \frac{c_{k,m,q_{k,m}} p_{k,m}}{N_0 + \sum_{q' \neq q_{k,m}} p_{k,q',m} c_{k,m,q'}} = \xi_{k,m} p_{k,m} \quad (4.3)$$

on $c_{k,m,q} = |\mathbf{h}_{k,m}^H \mathbf{b}_{q,m}|^2$ és el canal equivalent de l'usuari k , a la freqüència m del *beam* q —l'índex $q_{k,m}$ denota el *beam* assignat a l'usuari k de la freqüència m — i $\xi_{k,m}$ és la CINR².

Observant 4.2 s'aprecia que, en un instant donat, per cada freqüència es poden multiplexar fins a N_T senyals diferents espacialment. Per trobar aquest màxim, i treballant en un entorn de *broadcasting*, es suposa que els N_T transmissors —o l'estació base amb N_T antenes— coneixen tot el conjunt de $\gamma_{k,m}$. Es correspondria amb un sistema amb CSI parcial, ja que només es coneix les diferents SINR. La prioritat dels usuaris ve donada per ϕ_k i ha de satisfer la condició $\sum_{k=1}^K \phi_k = 1$ per donar consistència al sistema. Aquests paràmetres distribueixen la unitat del *sumrate* entre tots els usuaris i defineixen la proporció que s'assigna a cada usuari. Així, $0 < \phi_k \leq 1$, $\forall k = 1, \dots, K$.

²La CINR és la relació de potència del canal amb la de les interferències i el soroll. No es té present la potència transmesa.

De la primera restricció es desprèn la idea que tot el conjunt de les potències no pot ser superior a la potència màxima de la que disposa el sistema, però en valor mitjà i no instantani. La segona restricció reflecteix que la velocitat individual (part esquerra de la desigualtat) ha de ser igual o superior al *sumrate* ponderat per la seva prioritat. És a dir, ha de ser com a mínim la proporció que li pertoca.

Un dels avantatges de l'ús de l'optimització ergòdica en detriment de la instantània és la possibilitat d'assignar més potència del màxim permès si les condicions són bones perquè, en termes de mitjana, la potència assignada no superarà la màxima permesa. Així doncs, si en un instant donat les condicions són favorables, el sistema assignarà més potència per aprofitar-ho. D'altra banda, si les condicions canvien i es tornen hostils, s'assignarà menys potència. Però en termes de mitjana, la potència total que s'assignarà no serà superior al llindar establert.

Un altre avantatge de l'optimització ergòdica és la reducció de la complexitat de l'algoritme que en resulti. A més a més, la incorporació de la component temporal fa que es pugui explotar pels motius explicats anteriorment. L'assumpció d'ergodicitat del canal resulta un requisit primordial per mantenir l'optimització. A diferència d'altres algorismes instantanis, el que es proposa és incorporar la idea de *slot* temporal, que també s'adequa a la tecnologia actual.

L'assignació de potències, portadores, velocitats i *slots* temporals a diferents usuaris fa que la interacció entre les capes PHY i MAC sigui més estreta. L'accés es pot atorgar segons aquests paràmetres. La capa MAC s'encarrega de controlar la congestió de la xarxa reassignant les diferents prioritats o pesos segons les seves necessitats i la PHY, de gestionar l'accés en sí mateix. Existeixen nombroses tècniques per fer front a aquesta problemàtica però s'escapa de l'objectiu d'aquest projecte.

4.2 Optimizació dual

El problema plantejat no sempre és còncav i fa difícil trobar una solució que maximitzi el resultat. Això equival a afirmar que no sempre es pot obtenir una solució real i, per tant, no realitzable o senzillament una expressió tancada que doni resposta. No obstant, el problema es pot separar entre les freqüències per eliminar la dimensió freqüencial. Això es pot fer en OFDMA perquè totes les portadores són independents entre sí i, per tant, separables. Malgrat aquest benefici, cal tenir present que la restricció de potència i de velocitats és conjunta entre freqüències. És a dir, les restriccions inclouen la diversitat freqüencial. Per tant, tot i que la funció objectiu és separable freqüencialment no ho són així les restriccions. Tot això fa que el problema degeneri en no còncav.

Per resoldre'l es proposa utilitzar l'optimització dual. Aquesta consisteix en formular un problema diferent però que tingui els extrems en els mateixos punts que el problema inicial. D'aquesta manera, resolent el problema dual —de menys complexitat— s'obté una solució de punts que l'optimitzen i que, alhora, també són solució del problema inicial. Una manera d'implementar el problema dual és a partir dels multiplicadors de Lagrange. Aquest mètode inclou les restriccions dins de l'expressió a optimitzar i les elimina del problema. És a dir, deixen de ser restriccions en sí mateixes per ser part de l'equació a resoldre. D'aquesta manera es relaxa el problema perquè ja no hi ha tantes restriccions.

Per tant, el problema dual ara té com a variables els multiplicadors de Lagrange. Trobant una solució pels multiplicadors de Lagrange, es trobarà una solució al problema dual. És a dir, resolent un problema dual es resol el problema inicial.

Definint $R_{k,m} = \log_2(1 + \gamma_{k,m}(p_{k,m}))$ com la velocitat o *rate* de l'usuari k -èssim a la portadora m -èssima, el lagrangiana resulta

$$L(R, \mathbf{P}, \lambda, \boldsymbol{\mu}) = R + \lambda \left(\bar{P} - \mathbb{E} \left\{ \sum_{k=1}^K \sum_{m=1}^M p_{k,m} \right\} \right) + \sum_{k=1}^K \mu_k \left(\mathbb{E} \left\{ \sum_{m=1}^M R_{k,m} \right\} - \phi_k R \right). \quad (4.4)$$

El vector $\boldsymbol{\mu} = [\mu_1 \dots \mu_K]^T \in \mathbb{R}^K$ és el vector de multiplicadors que forcen la restricció de velocitats i λ , la de potència. La funció objectiu a partir de les variables λ i $\boldsymbol{\mu}$ es pot escriure

$$\begin{aligned} \Theta(\lambda, \boldsymbol{\mu}) &= \max_{R \in \mathbb{R}^+, \mathbf{P}} L(R, \mathbf{P}, \lambda, \boldsymbol{\mu}) \\ &= \lambda \bar{P} + \max_{R \in \mathbb{R}^+, \mathbf{P}} R(1 - \boldsymbol{\mu}^T \boldsymbol{\phi}) - \lambda \mathbb{E} \left\{ \sum_{k=1}^K \sum_{m=1}^M p_{k,m} \right\} + \sum_{k=1}^K \mu_k \mathbb{E} \left\{ \sum_{m=1}^M R_{k,m} \right\}. \end{aligned} \quad (4.5)$$

El vector $\boldsymbol{\phi} = [\phi_1 \dots \phi_K]^T$ correspon al vector de prioritats. Observant el primer operand, si $1 - \boldsymbol{\mu}^T \boldsymbol{\phi} > 0$, la solució òptima seria aquella on $R^* = \infty$. Resulta una solució inconsistent i no realitzable. Per l'altra banda, si $1 - \boldsymbol{\mu}^T \boldsymbol{\phi} < 0$ la solució seria $R^* = 0$. Que la velocitat sigui nul·la no és quelcom que sigui d'interès perquè el que es busca és precisament el contrari, que sigui màxima. Així doncs, el cas interessant és el que satisfà la condició $\boldsymbol{\mu}^T \boldsymbol{\phi} = 1$ i permet eliminar aquest operand ja que $1 - \boldsymbol{\mu}^T \boldsymbol{\phi} = 0$. Per tant, es pot simplificar per l'expressió

$$\Theta(\lambda, \boldsymbol{\mu}) = \lambda \bar{P} + \max_{\mathbf{P}} \sum_{k=1}^K \mu_k \mathbb{E} \left\{ \sum_{m=1}^M R_{k,m} \right\} - \lambda \mathbb{E} \left\{ \sum_{k=1}^K \sum_{m=1}^M p_{k,m} \right\}. \quad (4.6)$$

Observant que el problema s'ha descompost en suma de “subproblemes”, es pot reescriure l'anterior equació de la forma

$$\Theta(\lambda, \boldsymbol{\mu}) = \lambda \bar{P} + \sum_{k=1}^K \Theta_k(\lambda, \mu_k) \quad (4.7)$$

amb $\Theta_k(\lambda, \mu_k) = \max_{\mathbf{P}} \mu_k \mathbb{E} \left\{ \sum_{m=1}^M R_{k,m} \right\} - \lambda \mathbb{E} \left\{ \sum_{m=1}^M p_{k,m} \right\}$. Aquesta descomposició es pot analitzar com si fos la suma de optimitzacions per usuari. Així es pot distribuir la complexitat del problema entre els usuaris. Cada terminal resol o contribueix a resoldre el problema dual i l'estació base s'encarrega del conjunt i de resoldre la restricció de potència. Aquesta observació, feta a [HAN33], és clau pel desenvolupament del problema i que resulti viable la seva implementació.

La solució òptima per un k arbitrari es troba derivant els diferents $\Theta_k(\lambda, \mu_k)$ i s'obté l'expressió clàssica de *water-filling*

$$p_{k,m}^* = \left[\frac{\partial \Theta_k(\lambda, \mu_k)}{\partial p_{k,m}} = 0 \right] = \left[\frac{\mu_k}{\lambda \ln 2} - \frac{1}{\xi_{k,m}} \right]^+ \quad (4.8)$$

on $[x]^+ = \max(x, 0)$. Es posa de manifest que quan més alta sigui la SINR d'un usuari k en una portadora m , més potència se li assignarà.

Així doncs, el problema dual permet trobar els valors de λ i $\boldsymbol{\mu}$ de manera que facin tendir a zero la diferència entre la funció i la restricció. Per tant, es pot escriure el problema dual com

$$g^*(\lambda, \mu) = \min_{\lambda \geq 0, \mu \in \mathcal{D}} \Theta(\lambda, \mu) \quad (4.9)$$

on $\mathcal{D} = \{\mu \geq 0 : \mu^T \phi = 1\}$. Com s'aprecia, el problema inicial no assegurava la concavitat. En canvi, el problema dual sí presenta un mínim. És a dir, el problema dual és convex i, per tant, es pot trobar una solució per les variables duals que l'optimitzen i que alhora optimitzaran el problema inicial.

Tot i que és convexa pel conjunt de solucions de λ i μ , no es presenta diferenciable per tots els valors de λ . Així doncs, es proposa d'utilitzar un mètode de subgradient iteratiu per trobar la solució òptima. El mètode del subgradient es pot considerar una generalització del gradient matemàtic quan la funció que es manipula no és diferenciable en tots els punts. Aquest es fa de forma iterativa buscant una direcció perpendicular en cada moment per arribar al valor òptim.

Donat un λ^0 inicial, el subgradient respecte λ genera un conjunt de punts que, en una iteració i , són

$$\lambda^{i+1} = [\lambda^i - s_\lambda^i g_\lambda^i]^+ \quad (4.10)$$

on s_λ^i és un pas escalar i positiu i g_λ^i denota el subgradient de $\Theta(\lambda^*(\mu^i), \mu^i)$ respecte λ . De la mateixa manera, es pot utilitzar el mateix mètode per trobar μ^* que pren la forma

$$\mu^{i+1} = \Pi_{\mathcal{D}} [\mu^i - s_\mu^i g_\mu^i] \quad (4.11)$$

on s_μ^i és un pas escalar i positiu i g_μ^i denota el subgradient de $\Theta(\lambda^*(\mu^i), \mu^i)$ respecte μ i $\Pi_{\mathcal{D}}$ denota la projecció al subespai \mathcal{D} .

Com es veu, el subgradient troba la solució al problema dual perquè la restricció de potència ja és en sí mateixa el subgradient si s'evalua al punt òptim de potència \mathbf{P}^* . Així es té

$$g_\lambda^i = \bar{P} - \mathbb{E} \left\{ \sum_{k=1}^K \sum_{m=1}^M p_{k,m}^*(\lambda^i, \mu_k^i) \right\} \quad (4.12)$$

per uns λ^i i μ^i donats; i

$$g_\mu^i = \bar{\mathbf{R}}^i - \phi \bar{R}^i \quad (4.13)$$

on $\bar{\mathbf{R}}^i = [\bar{R}_1^i \dots \bar{R}_K^i]^T$ amb $\bar{R}_k^i = \mathbb{E} \left\{ \sum_{m=1}^M R_{k,m} \right\}$ és el vector de les velocitats per cada usuari, i

$$\bar{R}^i = \sum_{k=1}^K \bar{R}_k^i = \mathbb{E} \left\{ \sum_{k=1}^K \sum_{m=1}^M R_{k,m} \right\} \quad (4.14)$$

és el *sumrate*. La projecció al subespai \mathcal{D} que satisfà la condició $\mu^T \phi = 1$ es realitza reescalant el vector de la forma

$$\Pi_{\mathcal{D}}[\mu] = \frac{[\mu]^+}{\phi^T [\mu]^+}. \quad (4.15)$$

El vector μ es pot interpretar com la restricció que ha de seguir cada usuari per garantir que la seva velocitat mitjana s'assoleix i que queda garantitzada. Donat que λ depèn en sí



Figura 4.1: Diagrama de blocs de tot el procés

mateixa del valor de μ , un cop μ convergeix, λ també ho fa i s'acompleix la restricció de potència.

Com s'ha vist en aquest punt, el mètode de subgradient dona resposta al problema dual i, a més, presenta una forma senzilla i gens complexa d'implementació que tot seguit es veurà.

4.3 Algoritme

De la solució que es desprèn de 4.8 s'aprecia que la potència es va distribuint entre els usuaris i portadores segons la tècnica de *water-filling*. Aquesta tècnica “activa” les portadores als usuaris però no els assigna un accés no interferent. Així doncs, per una mateixa portadora es pot donar el cas que varis usuaris tinguin potència assignada. Aquesta assignació no resol, però, el problema d'interferència. Indica que en una portadora han d'anar multiplexats varis usuaris però no mostra com. El multiplexat no es pot realitzar en temps, ja que donat que l'objectiu del problema és en termes ergòdics no es pot utilitzar la component temporal. El que fa, per tant, és assignar accés de forma freqüencial. Cal, llavors, multiplexar-los d'una altra manera. La figura 4.1 il·lustra l'esquema que es seguirà per implementar aquest algoritme.

4.3.1 Planificador espacial

El multiplexat espacial dona resposta aquest problema. Partint de que la planificació freqüencial proposada assignarà accés a múltiples usuaris en una sola portadora, cal trobar prèviament quins usuaris disposaran d'un millor canal. En essència es vol escollir quins usuaris són els més òptims per transmetre portadora a portadora. Del punt de vista del *cross-laying*, en surt un gestor espacial d'accés. De [TAE04] es desprèn la idea de multiplexar aquells usuaris que disposin d'un canal més favorable. Per això, és necessari sondar quins canals són i en quines condicions ho són.

Tot i proposar un planificador espacial, la manera més òptima de resoldre el problema seria trobar una solució conjunta al problema. És a dir, resoldre la planificació freqüencial i espacial alhora. Una manera de fer-ho és a través de funcions útils. Tal i com proposa [UTS23], la proposició d'aquestes funcions fa que es trobi una solució directa optimitzant totes les variables alhora. No obstant, la complexitat que presenta és significativa. Per això, malgrat sacrificar una mica d'“optimibilitat”, l'esquema de conformació oportunista [HAS34] presenta una bona resposta, sobretot amb un alt nombre d'usuaris, i una baixíssima complexitat de càlcul.

Per iniciar l'esquema oportunista, es generen fins a N_T *beams* ortonormals i aleatoris i s'escullen els usuaris que més bona qualitat de senyal reben. Degut a la seva aleatorietat, farà que en un instant es multiplexin uns usuaris i en un instant, uns altres. S'ha de tenir present

que aquesta tècnica produeix interferències entre els usuaris. És a dir, cada usuari percebrà una interferència generada per la resta de *beams*. Per tant, si es busca maximitzar la velocitat total del sistema, pot donar-se el cas que no sempre s'hagin de generar N_T *beams*.

Per realitzar aquest sondatge s'empregarà el mateix nivell de potència a tots els usuaris, *beams* i portadores. Això equival a $p_{k,m,q} = P_s, \forall k, m, q$. D'aquesta manera es realitza una estimació a priori del valor que prendrà γ . En una portadora m , es defineix $Q_m \in \{1, \dots, N_T\}$ com el nombre d'usuaris multiplexats espacialment a la portadora m . Donat que, a priori, no hi ha assignació espacial, la SINR d'un usuari k a la portadora m pel *beam* q és

$$\hat{\gamma}_{k,m,q}^{(Q_m)} = \frac{c_{k,m,q}}{\frac{N_0}{P_s} + \sum_{q' \neq q}^{N_T} c_{k,m,q'}} = \hat{\xi}_{k,m,q}^{(Q_m)} P_s. \quad (4.16)$$

Tenint en compte que Q_m és una variable (posteriorment a optimitzar), existeixen $\binom{K}{Q_m} Q_m!$ permutacions possibles de $\hat{\gamma}_{k,m,q}^{(Q_m)}$. Definint el conjunt $S_m^{(Q_m)} \subseteq \{1, \dots, N_T\}$ com el conjunt de *beams* i $U_m^{(Q_m)} \subset \{1, \dots, K\}$ com el conjunt d'usuaris a la portadora m i amb Q_m elements cadascun sense repetició, els conjunts òptims S_m^* i U_m^* que maximitzen el *sumrate* es troben realitzant recerques exhaustives. No obstant, es pot realitzar una aproximació per reduir la complexitat encara que la solució sigui subòptima. Definint l'índex $j = 1, \dots, \binom{N_T}{Q_m}$ com a l'índex de la permutació del conjunt $S_m^{(Q_m)}$, si es fixa aquest índex a un j o permutació arbitraris, llavors l'usuari òptim satisfà la condició

$$k_{m,j,q}^* = \arg \max_k \hat{\gamma}_{k,m,j,q}^{(Q_m)}. \quad (4.17)$$

Aquest usuari és agregat al conjunt $U_{j,m}^{(Q_m)}$. Anàlogament, s'afegeix el *beam* q al conjunt $S_{j,m}^{(Q_m)}$. És important remarcar que en aquest punt es crea una relació unívoca entre usuari k i *beam* q . És a dir, $k \leftrightarrow q$. Parlar d'un usuari k equival a relacionar-lo amb un sol *beam* q . Noti's que aquesta relació està definida per cada permutació j -èsima.

Per cada freqüència, la permutació òptima es troba a partir de

$$j_m^* = \arg \max_{1 \leq j \leq \binom{N_T}{Q_m}} \sum_{q=1}^{Q_m} \log_2 \left(1 + \hat{\gamma}_{k_{m,j,q}^*, m, j, q}^{(Q_m)} \right). \quad (4.18)$$

Per tant, el conjunt òptim d'usuaris que en serveix Q_m és $U_m^{*(Q_m)} = U_{j_m^*, m}^{(Q_m)}$ i anàlogament, pel conjunt dels *beams* $S_m^{*(Q_m)} = S_{j_m^*, m}^{(Q_m)}$.

Finalment, el nombre d'usuaris servits òptims es troba a partir de

$$Q_m^* = \arg \max_{1 \leq Q_m \leq N_T} \sum_{k \in U_m^{*(Q_m)}} \sum_{q \in S_m^{*(Q_m)}} \log_2 \left(1 + \hat{\gamma}_{k, m, j_m^*, q}^{(Q_m)} \right). \quad (4.19)$$

En aquest punt, a cada portadora s'assigna un conjunt d'usuaris i *beams* corresponents. Es té un multiplexat espacial. Així doncs, unit amb la solució anterior de multiplexat freqüencial es poden assignar accessos a múltiples usuaris en múltiples freqüències i múltiples direccions.

Un dels avantatges d'aquesta tècnica és que es poden utilitzar altres codificadors espacials, com poden ser el forçador de zeros o diagonalització del canal, encara que incrementen la complexitat sobretot en el canal de *feedback*. Tot i haver proposat l'esquema oportunista per poca complexitat —es recorda que és un dels objectius que es persegueix en aquest projecte—, no és condició indispensable pel seu funcionament.

4.3.2 Planificador freqüencial

Un cop realitzada una planificació espacial, es pot procedir amb l'assignació de potència per portadora i usuari. Per implementar la solució iterativa del problema dual es proposa utilitzar l'aproximació estocàstica de [WON35] que permet dur una assignació adaptativa dels recursos OFDMA sense haver de conèixer la distribució estadística del canal. És a dir, no és necessari conèixer quina estadística segueix el canal perquè només són necessaris els valors instantanis de la SINR.

L'aproximació estocàstica es basa en el principi que tota observació d'un paràmetre es pot anar aproximant a partir de l'observació d'una variable aleatòria i un error. Així, definint $\theta[n]$ com el paràmetre a observar a la iteració n , es té

$$\theta[n+1] = \theta[n] + \epsilon[n]\Psi[n] \quad (4.20)$$

on $\Psi[n]$ és la variable aleatòria i $\epsilon[n] > 0$ és un error o pas que es pot anar decrementant amb les iteracions. Es pot provar que a mesura que s'itera convergeix a un punt estacionari $\lim_{n \rightarrow \infty} \theta[n] = \theta^*$.

Aquesta idea és la que s'utilitza per trobar els valors de λ i μ . A partir de les iteracions de 4.10 i 4.11 del subgradient, es pot utilitzar l'aproximació estocàstica imposant que les iteracions formin part de la component temporal. És a dir, efectuar cada iteració en diferents instants o intervals de temps. Així es pot reescriure 4.10 com

$$\lambda[n+1] = [\lambda[n] + \beta[n]g_\lambda[n]]_\epsilon^+ \quad (4.21)$$

i 4.11

$$\mu[n+1] = \Pi_{\mathcal{D}} [\mu[n] - \beta[n]g_\mu[n]] \quad (4.22)$$

amb $[x]_\epsilon^+ = \max(x, \epsilon)$ on $0 < \epsilon \ll 1$ és una constant. $\beta[n]$ és el pas que dona més o menys importància al subgradient i ha de complir la condició

$$\sum_{n=0}^{\infty} \beta[n] = \infty, \quad \beta \geq 0, \quad \beta \rightarrow 0 \quad (4.23)$$

per satisfer la convergència.

Un cop realitzat el multiplexat espacial, fer referència a $\gamma_{k,m,q}$ és equivalent a $\gamma_{k,m}$ donat que existeix una relació unívoca entre *beam* i usuari per una portadora concreta. Així doncs, aplicant el mateix raonament, la potència assignada és

$$\tilde{p}_{k,m}[n] = \left[\frac{\mu_k[n]}{\lambda[n] \ln 2} - \frac{P_s}{\hat{\gamma}_{k,m}[n]} \right]^+ \quad (4.24)$$

$$p_{k,m}^*[n] = \begin{cases} \tilde{p}_{k,m}[n], & k \in U_m^* \\ 0, & k \notin U_m^* \end{cases} \quad (4.25)$$

on $\hat{\gamma}_{k,m}[n]$ és la SINR de l'usuari k a la portadora m a l'instant n . Cal tenir molt present que $\hat{\gamma}_{k,m}[n]$ no és en sí la SINR real, sinó que correspon a l'expressió 4.16, que és la que han reenviat els usuaris i que es realitza amb $p_{k,m} = P_s$. Donat que hi ha una dependència mútua (ambdues són funció d'ambdues, $\gamma(\mathbf{P})$ i $\mathbf{P}(\gamma)$), l'equació 4.16 trenca aquesta dependència tot i produir un resultat subòptim, ja que el paràmetre d'entrada SINR és una estimació.

La velocitat instantània per usuari és

$$R_k[n] = \sum_{m=1}^M R_{k,m}[n] = \sum_{m=1}^M \log_2 \left(1 + \hat{\gamma}_{k,m} \frac{p_{k,m}^*}{P_s} \right) \quad (4.26)$$

i la potència instantània total del sistema,

$$P[n] = \sum_{k=1}^K \sum_{m=1}^M p_{k,m}^*. \quad (4.27)$$

Les aproximacions del subgradient resulten

$$g_\lambda[n] = \bar{P} - P[n] \quad (4.28)$$

$$\mathbf{g}_\mu[n] = \mathbf{R}[n] - \phi R[n] \quad (4.29)$$

on $\mathbf{R}[n] = [R_1[n] \dots R_K[n]]^T$ és el vector de velocitats i $R[n] = \sum_{k=1}^K R_k[n]$ és el *sumrate* del sistema.

Tal com s'ha vist, la solució de les variables duals es realitza de forma adaptativa a partir del subgradient per aproximació estocàstica.

4.4 Convergència

S'entén com a convergència el fet que l'algoritme convergeixi realment cap a la solució òptima i no altra. És a dir, que l'assignació de potència sigui realment l'òptima. D'entrada ja s'afegeix que aquest fet ve contrarestat per l'ús de l'estratègia oportunista, l'ús de la potència de sondeig P_s per estimar la SINR i la configuració dels paràmetres (Q_m fixa o variable) del planificador espacial. Tots aquests punts contribueixen directament al grau d'optimització. No obstant, deixant de banda aquests factors insalvables, es vol estudiar si el resultat que s'obté es correspon amb l'òptim teòric.

La convergència d'aquest algoritme presenta un punt crític ja que està dividit en dues grans etapes. La planificació espacial i freqüencial es duu a terme en dos blocs separats i això fa que es produeixi un desajustament a priori del grau d'optimització. No obstant, gràcies a la maximització ergòdica, es va reajustant als valors òptims. L'optimització ergòdica resulta molt útil en aquest àmbit perquè permet que, en el transcurs del temps, l'algoritme es vagi adaptant i disminuint els errors que es cometien i que desfavoreixen el resultat.

Així doncs, es podria estudiar la convergència a partir de la convergència per separat dels dos blocs. La convergència del primer bloc, el planificador espacial, no presenta un problema en sí mateixa perquè no representa cap problema a optimitzar. Senzillament es fa una ordenació espacial dels millors usuaris, es seleccionen i es passen al següent bloc. Així, aquesta part realitza la seva funció, amb més o menys complexitat, però sempre ofereix un conjunt d'usuaris que obtenen les millors prestacions en termes de velocitat. Per tant, tot i ser òptim en la selecció d'usuaris (suposant que es deixen tots els paràmetres com a variables), no ho és el conformador. Donat que es tracta de l'estratègia oportunista, aquesta en sí mateixa no és òptima. El fet de generar *beams* aleatòriament fa que no sigui plenament òptima. Com s'ha vist, el grau d'optimització ve donat, principalment, pel nombre d'usuaris. Així doncs, la convergència d'aquest bloc presenta una solució subòptima. És important afegir que l'assignació de potència amb diversitat espacial pel canal de *downlink* és encara un problema obert, com es mostra a [ONE36].

Pel que fa al segon bloc, la convergència que cal analitzar en detall correspon a la del planificador freqüencial. L'optimització dual és per si mateixa una eina que garanteix que el resultat que s'obté és realment l'òptim, tot i resoldre's en un problema dual. No obstant, existeix el que s'anomena *duality gap*. Aquest és la diferència entre la solució del problema inicial i la del dual. L'apèndix A.5 es dedica a estudiar aquest aspecte de l'optimització i ofereix una condició per garantir el grau d'optimització. A partir d'aquest punt cal tenir-se present la demostració d'aquest apèndix.

Per demostrar que l'algoritme d'aquest capítol convergeix a la solució òptima cal demostrar que és consistent amb les condicions i el teorema de l'apèndix A.5. Primerament, per ser OFDM, ja es pot afirmar que el sistema compleix la propietat de *time-sharing* i que, per tant, el *duality gap* és nul. Cal tenir present que aquest fet es produeix estrictament quan $M \rightarrow \infty$. En el cas que ocupa aquest projecte, es suposarà menyspreable.

Seguidament, només cal veure que realment es compleix la formulació d'aquest apèndix. Per fer-ho es fan les següents assignacions

$$\begin{aligned}
 f_m(x_m) &= \mathbb{E} \left\{ \sum_{k=1}^K \log_2 (1 + \gamma_{k,m}(\mathbf{P})) \right\} \\
 h_m(x_m) &= \begin{bmatrix} \mathbb{E} \left\{ \sum_{k=1}^K p_{k,m} \right\} \\ \mathbb{E} \{ \log_2 (1 + \gamma_{1,m}(\mathbf{P})) \} \\ \vdots \\ \mathbb{E} \{ \log_2 (1 + \gamma_{K,m}(\mathbf{P})) \} \end{bmatrix} \\
 \mathbf{\Lambda} &= \begin{bmatrix} \bar{P} \\ \phi_1 R \\ \vdots \\ \phi_K R \end{bmatrix}.
 \end{aligned} \tag{4.30}$$

Per tant, queda posat de manifest que l'algoritme plantejat compleix les propietats que fan que el *duality gap* sigui nul. D'aquesta manera, la solució que ofereix es correspon amb l'òptima.

4.5 Estabilitat

L'estabilitat d'aquest algoritme ve determinada, sobretot, pel planificador freqüencial. Donat que el planificador espacial sempre realitza la seva tasca instantàniament, multiplexant a cada moment el conjunt d'usuaris més òptim, l'estudi de l'estabilitat recau principalment sobre la gestió freqüencial.

Per fer-ho s'analitza la variància de g_λ . Per això, desenvolupant les expressions es pot escriure

$$\begin{aligned}
\sum_{n=0}^{\infty} \mathbb{E} \{ \beta[n]^2 |g_\lambda[n]|^2 \} &= \sum_{n=0}^{\infty} \beta[n]^2 \mathbb{E} \{ (\bar{P} - P[n])^2 \} \\
&= \sum_{n=0}^{\infty} \beta[n]^2 (\bar{P}^2 - \bar{P} \mathbb{E} \{ P[n] \} + \mathbb{E} \{ P[n]^2 \}) \\
&\leq \bar{P}^2 \sum_{n=0}^{\infty} \beta[n]^2 + \sum_{n=0}^{\infty} \beta[n]^2 \mathbb{E} \{ P[n]^2 \}.
\end{aligned} \tag{4.31}$$

El primer sumand queda limitat clarament per l'elecció del pas $\beta[n]$. Per tal de que quedi fitat superiorment, ha de complir unes condicions. Per exemple, si pren la forma

$$\beta[n] \propto \frac{1}{n} \tag{4.32}$$

llavors complirà les condicions 4.23 i, a més, $\sum_{n=0}^{\infty} \beta[n]^2 = 0$.

A més,

$$\begin{aligned}
\mathbb{E} \{ P[n]^2 \} &= \mathbb{E} \left\{ \left(\sum_{k=1}^K \sum_{m=1}^M p_{k,m}^*[n] \right)^2 \right\} \\
&\leq (KM)^2 \mathbb{E} \left\{ \left(\max_{k,m} \tilde{p}_{k,m}[n] \right)^2 \right\} \\
&\leq (KM)^2 \left(\max_k \frac{\mu_k[n]}{\ln 2 \lambda[n]} \right)^2
\end{aligned} \tag{4.33}$$

està fitat superiorment ja que $\mu_k[n] \leq B_\mu < \infty$, $\forall n$ i $\lambda[n] \geq \epsilon$, $\forall n$. Així doncs, es té

$$\begin{aligned}
\sum_{n=0}^{\infty} \mathbb{E} \{ \beta[n]^2 |g_\lambda[n]|^2 \} &\leq \bar{P}^2 \sum_{n=0}^{\infty} \beta[n]^2 \mathbb{E} \{ P[n]^2 \} \\
&\leq \bar{P}^2 \sum_{n=0}^{\infty} \beta[n]^2 + \left(KM \frac{B_\mu}{\epsilon \ln 2} \right)^2 \sum_{n=0}^{\infty} \beta[n]^2 \\
&< \infty.
\end{aligned} \tag{4.34}$$

Per tant, queda demostrat que la variància del subgradient respecte la potència queda fitada i que sempre prendrà valors reals dins d'un marge.

4.6 Complexitat

Un dels principals objectius plantejats amb aquest problema és que la complexitat sigui relativament acceptable. Les solucions actuals certament són més òptimes però en detriment d'una complexitat baixa. Es requereixen grans càlculs operacionals per optimitzar poc el problema. Així, es planteja el compromís entre complexitat i quant òptim es vol que es sigui.

Per això és important marcar la proximitat de les cotes “ideals” i avaluar aquesta complexitat. Si els resultats són pròxims als òptims i ho fan amb baixa complexitat, és evident que sí és bo tenir-ho en compte.

La complexitat d'aquest algoritme ve diferenciada principalment per la complexitat dels dos blocs. L'ús de l'ergodicitat efectivament redueix la complexitat del pla freqüencial però cal analitzar el pla espacial.

Tot el procés queda sintetitzat amb la figura 4.1. En el primer bloc, l'estació base genera N_T *beams* ortonormals i va activant-los un per un. Cada usuari reté el nivell de SINR. La complexitat en aquest procés és de $\mathcal{O}(MN_T)$.

El següent pas és el còmput de totes les permutacions possibles de $\hat{\gamma}$. Això equival a realitzar tots els càlculs de $\hat{\gamma}_{k,m,j,q}^{(Q_m)}$. La complexitat d'aquest procés és, com es pot veure, binomial. Així, es té $\mathcal{O}(MK2^{N_T}N_T^2)$.

Seguidament, el pla freqüencial disposa de poca complexitat. Així, durant el procés de *water-filling*, només es té $\mathcal{O}(MK)$ i, finalment, l'actualització dels paràmetres només requereix $\mathcal{O}(M)$.

Així doncs, es pot apreciar que el pla freqüencial és de baixa complexitat però no així el pla espacial. No obstant, aquesta complexitat pot quedar dràsticament reduïda.

4.6.1 Complexitat reduïda

L'ús de Q_m com a variable en el planificador espacial fa que el sistema resulti molt més òptim però per contra també fa augmentar la complexitat. Deixant Q_m com a variable, fa que el nombre d'operacions segueixi una corba binòmica. Aquest paràmetre descriu el nombre òptim d'usuaris a servir en cada instant. No obstant, s'observa que Q_m roman força estàtica en tot el procés. És a dir, té tendència a oscil·lar al voltant d'un valor i varia amb poca freqüència. Així doncs, es fa fàcil pensar de deixar aquest paràmetre com a fix passades unes quantes iteracions.

Si el paràmetre Q_m es deixa fix a un valor, la complexitat es veu reduïda —també queda reduït el *feedback*, tal i com es veurà a continuació—. Així, amb el paràmetre fix, la complexitat resulta $\mathcal{O}(MK2^{N_T}N_T)$. Tot i que ja no és quadràtica, continua essent binomial. No obstant, si es fixa $Q_m = Q = N_T$ al màxim d'antenes i per totes les freqüències, el nombre de permutacions binomials esdevé la unitat, ja que $\binom{N_T}{N_T} = 1$. Així doncs, amb aquesta condició la complexitat es redueix fins a $\mathcal{O}(MKN_T)$. No obstant, els resultats que s'obtingran seran subòptims, perquè es perd un grau de llibertat i, per tant, té menys capacitat per adaptar-se al medi. Tota la complexitat queda resumida a la taula 4.1.

4.7 Feedback

El *feedback* o la informació que han de remetre els usuaris a l'estació base és una component essencial del sistema. Quan l'estació sondeja els diferents canals amb els diferents *beams*, els usuaris han de reenviar els valors de les diferents $\hat{\gamma}$ que perceben. És molt important remarcar que no es reenvia tot el canal \mathbf{h} , sinó que només es reenvia el valor de senyal percebut. En altres paraules, es tracta de CSI parcial perquè no és necessari disposar de tota la informació de canal. El nombre de paràmetres a reenviar és important perquè el canal de *feedback* és molta menys capacitat que el principal degut a que els terminals han d'utilitzar la menor potència possible per fer durar les bateries. A més, cal tenir present que segons si Q_m es deixa fixa o variable, el nombre de paràmetres canviarà i serà diferent.

Taula 4.1: Algoritme pas a pas amb la complexitat

1. Sondeig: l'estació base transmet un senyal pilot per sondejar cada canal equivalent	$\mathcal{O}(MN_T)$
2.a Q_m variable	
<i>Feedback.</i> Cada usuari envia $N_T x M$ paràmetres: $\left \mathbf{h}_{k,m}^T \mathbf{b}_{m,q} \right ^2$	$\mathcal{O}(MN_T)$
Planificació espacial	$\mathcal{O}(MK2^{N_T} N_T^2)$
2.b Q_m fix	
<i>Feedback.</i> Cada usuari envia $3xM$ paràmetres: $q_m^*, j_m^*, \hat{\gamma}_{k,m,j_m^*,q_m^*}$	$\mathcal{O}(M)$
Planificació espacial	$\mathcal{O}(MK2^{N_T} N_T)$
2.c $Q_m = N_T$	
<i>Feedback.</i> Cada usuari envia $2xM$ paràmetres: $q_m^*, \hat{\gamma}_{k,m,1,q_m^*}$	$\mathcal{O}(M)$
Planificació espacial	$\mathcal{O}(MKN_T)$
3. <i>Water-filling</i> freqüencial	$\mathcal{O}(MK)$
4. Actualització de paràmetres λ, μ	$\mathcal{O}(K)$

Deixant Q_m variable, un usuari k ha de reenviar tots els valors de $c_{k,m,q}$, per $m = 1, \dots, M$ i $q = 1, \dots, N_T$. Així, cada usuari no ha de realitzar cap càlcul ni cap estimació del canal $\mathbf{h}_{k,m}$ sinó que és suficient en mesurar el nivell de potència rebut.

Amb Q_m fixa, els usuaris han de conèixer prèviament aquest valor. Com que cada usuari disposa de la informació suficient donat que Q_m és fixa, cada usuari pot calcular les diferents permutacions de $\hat{\gamma}_{k,m,j,q}$ seguint la fórmula 4.16 per totes les permutacions de j . Un cop totes calculades, troba el valor més alt de $\hat{\gamma}_{k,m,j,q}$ i el reenvia, juntament amb els seus respectius índexs j i q . És a dir, l'usuari envia a l'estació quina permutació és la més bona i per quin *beam* prefereix el senyal, i el nivell de potència que percep pels anteriors índexs.

Finalment, de l'altre mode fix $Q_m = N_T$, donat que no hi ha permutacions possibles ($j = 1$ sempre), un usuari només cal que trobi el valor de senyal més alt i el *beam* corresponent.

Cal tenir present que aquest procés s'efectua per cada portadora, $m = 1, \dots, M$.

La reducció de *feedback* és un tema força recurrent a la investigació, donat la seva importància. Existeixen nombroses tècniques per OFDMA que redueixen la seva complexitat i la quantitat de paràmetres a enviar. Mètodes com l'agrupament freqüencial o el processat per *slots* es poden aplicar per reduir encara més el *feedback*.

La taula 4.1 il·lustra pas a pas l'algoritme proposat, amb la complexitat de cada fase. A més, també mostra els paràmetres que s'han d'enviar segons el mode Q_m variable o fix.

4.8 Fairness

La justícia en aquests àmbits és un terme força difús. Definir si un sistema és just o és injust resulta força subjectiu. En sí mateix implica que tots els usuaris reben el mateix tracte, en termes de velocitat, o no. No obstant, el cas presentat es desvia lleugerament d'aquest concepte ja que introdueix pesos o prioritats per regular la velocitat i obtenir una qualitat de servei concreta. Per tant, pel sistema presentat la justícia ve atorgada pel compliment de les prestacions: que els usuaris assoleixin la prioritat assignada.

El principal problema que presenta és que la selecció d'usuaris prèvia es realitza a partir del guany dels canals de tots els usuaris. Així es fa fàcil pensar que els usuaris amb millor canal

aconseguiran molt més accés que els altres. Per exemple, els usuaris més pròxims a l'estació base tindran poca atenuació i tendiran a monopolitzar els accessos. I si hi ha usuaris llunyans amb una prioritat important requeriran molts més accessos per aconseguir la QoS definida però que difícilment obtindran perquè tindran poc guany de canal. Tot això compromet el comportament de l'optimització dual.

Observant els multiplicadors μ s'aprecia que es van adaptant a les necessitats i proporcionalment a la prioritat. Així doncs, una manera de millorar el planificador espacial anterior és maximitzar la solució

$$\max_k \sum_{k=1}^K \mu_k \log_2 (1 + \hat{\gamma}_{k,m}), \quad \forall m. \quad (4.35)$$

Per tant, les expressions 4.18 i 4.19 queden modificades de forma

$$j_m^* = \arg \max_{1 \leq j \leq \binom{N_T}{Q_m}} \sum_{q=1}^{Q_m} \mu_k \log_2 \left(1 + \hat{\gamma}_{k_m^*, j, q, m, j, q}^{(Q_m)} \right) \quad (4.36)$$

i

$$Q_m^* = \arg \max_{1 \leq Q_m \leq N_T} \sum_{k \in U_m^*(Q_m)} \sum_{q \in S_m^*(Q_m)} \mu_k \log_2 \left(1 + \hat{\gamma}_{k, m, j_m^*, q}^{(Q_m)} \right) \quad (4.37)$$

respectivament.

Així s'assegura que tots els usuaris rebin l'accés adientment segons la seva prioritat assignada sense monopolis dels usuaris amb millor canal.

4.9 Resum

En aquest capítol s'ha vist l'objectiu que es vol aconseguir, el seu plantejament matemàtic, la solució i, finalment, la seva implementació. Pas a pas s'ha anat desgranant la idea principal i analitzant altres aspectes de conformació, com la complexitat o la convergència. El següent capítol està dedicat a plasmar amb resultats tota aquesta implementació i comentar aquells aspectes més atractius i importants.

Aquest capítol s'ha dedicat a presentar la solució a un problema mitjançant una optimització dual i de poca complexitat. El resultat és un algoritme que assigna potència en espai i freqüència explotant la diversitat temporal i d'usuari. Per això s'ha utilitzat una maximització ergòdica en detriment de la instantània.

Els avantatges queden sintetitzats a continuació:

- Baixa complexitat. L'algoritme és fàcil d'implementar i reparteix la complexitat entre els usuaris.
- Explotació de la diversitat temporal i reducció de la complexitat mitjançant la maximització ergòdica.
- Ofereix múltiples estratègies segons la configuració dels paràmetres que permeten més o menys complexitat a canvi de menors o majors resultats. Així es pot variar el grau d'optimització segons la potència de càlcul de la que es disposi.
- L'ús de prioritats o pesos fa que es pugui incorporar de manera senzilla prioritats de velocitat o varies QoS.

- Requereix només CSIT parcial i no és necessari conèixer quina estadística segueixen els diferents canals, ja que utilitza valors instantanis.
- Presenta poc *feedback*. Tot i això, es poden utilitzar altres estratègies per reduir-lo com l'agrupació de freqüències.
- Permet una quasi immediata incorporació a escenaris on utilitzin OFDMA, amb l'avantatge que això suposa. A més, no trenca esquemes existents com UMTS o HSDPA.
- No requereix que els usuaris disposin múltiples antenes. Per tant, abarateix costos en els terminals.

El pròxim capítol dedica a analitzar els resultats obtinguts a partir de simulacions realitzades. Amb aquests resultats es vol demostrar que aquest algoritme fa la seva tasca i que el que s'obté és atractiu. A més, també es vol posar de manifest la baixa complexitat que requereix. Aquests resultats mostren la resposta d'aquest algoritme en diferents situacions, com poden ser l'entrada i sortida d'usuaris, el canvi de prioritats, entre d'altres.

5 Resultats

EN aquest capítol s'exposaran els resultats obtinguts a partir de l'algoritme del capítol anterior. Per obtenir resultats sobre l'algoritme implementat s'han realitzat un conjunt de simulacions, variant els paràmetres de configuració i analitzant les diferents respostes que es donen. S'ha revisat, en primer lloc, que el que es proposa realment permet obtenir els resultats esperats. En segon lloc, s'ha comprovat que el que s'obté és consistent i estable. I finalment, s'ha comparat amb altres solucions més i menys òptimes i complexes.

Per donar més fonaments als resultats, s'han realitzat amb dos simuladors independents. D'aquesta manera s'ha comprovat que la implementació es pot portar d'un simulador a un altre sense incidents.

5.1 Escenari senzill

L'escenari simulat amb el que s'ha implementat l'algoritme contempla només esvaïments ràpids seguint l'estadística Rayleigh i sense LOS. No té pèrdues d'atenuació i sense esvaïments lents. És a dir, tots els terminals estan situats a la mateixa distància i equispaiats. Els paràmetres OFDMA són de $M = 64$ portadores, d'1 Hertz d'ample de banda. L'estació base disposa de $N_T = 2$ en ULA amb la distància entre antenes de $d = 0.5\lambda$. La potència mitja del sistema és de $\bar{P} = 10dB$ i la potència de sondeig, $P_s = 0dB$. Els paràmetres inicials de l'algoritme són $\lambda[0] = 3$, $\mu[0] = \frac{\phi}{\phi^T \phi}$, $\beta[n] = 0.01$. L'algoritme s'ha etiquetat com a EMOB (Enhanced Multiuser Opportunistic Beamforming).

5.1.1 Actuació

Primerament es vol observar que el sistema compleix les expectatives. La figura 5.1 mostra un escenari amb $K = 5$ usuaris amb prioritats diferents. El vector de prioritats és $\phi = [0.10 \ 0.15 \ 0.20 \ 0.25 \ 0.30]^T$. Per reflectir-ho amb més detall, la gràfica s'ha normalitzat. És a dir, en cada iteració, la velocitat s'ha dividit per la suma de tots. Així, es compleix la restricció de 4.2 de les velocitats de manera que

$$\frac{\bar{R}_k}{R} = \phi_k. \quad (5.1)$$

Com s'aprecia, en les primeres iteracions hi ha una lleugera adaptació. A més, les corbes estan amitjanades per estudiar el seu comportament. El *sumrate* mitjà aconseguit és de 360bps.

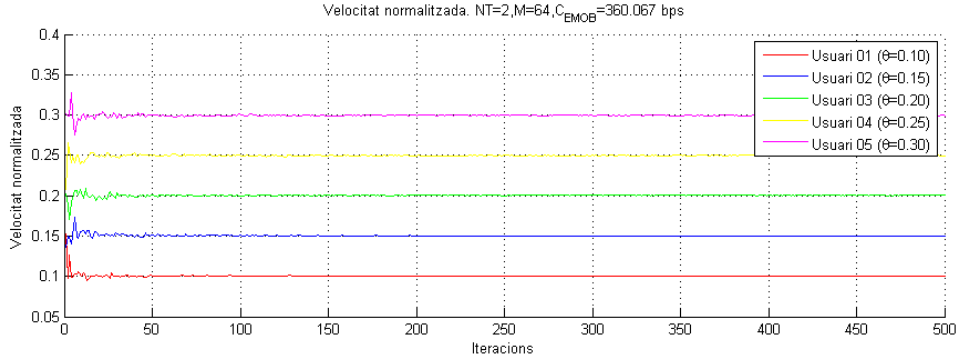
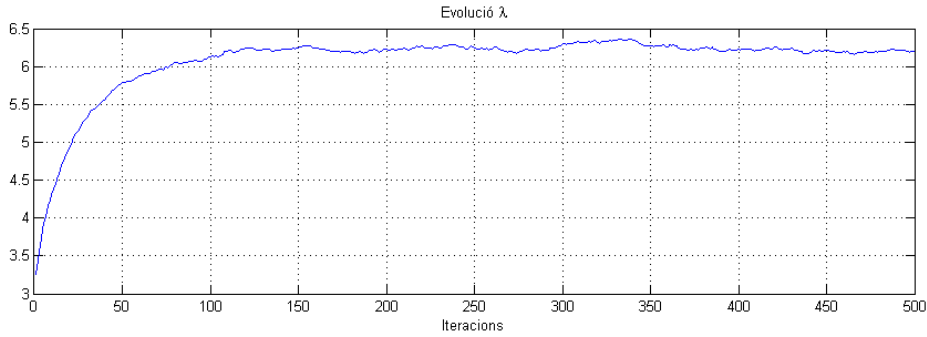
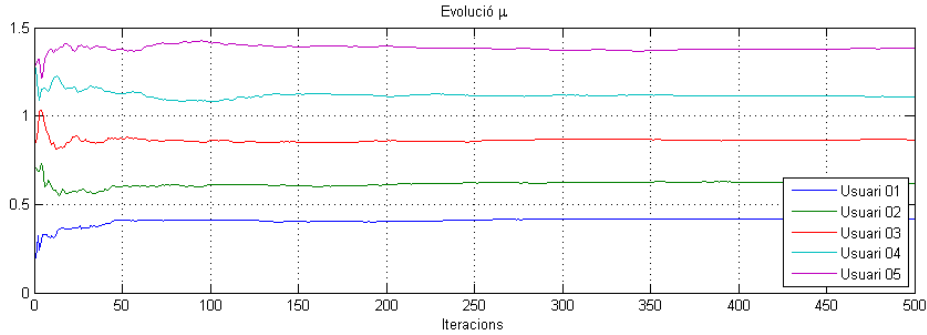


Figura 5.1: Velocitat mitja normalitzada amb 5 usuaris.

Figura 5.2: Evolució instantània de la λ .Figura 5.3: Evolució amitjanada de la μ .

Les cinc corbes convergeixen en poques iteracions al valor ϕ_k que se'ls ha assignat. Així doncs, per exemple, l'usuari 3 aconsegueix la cinquena part del *sumrate*, és a dir *72bps*.

Pel mateix cas, els paràmetres també oscil·len al voltant d'un valor mig. Tant per la λ com per la μ , ambdós convergeixen cap a un valor que garanteix les restriccions de potència i de velocitat. Les figures 5.2 i 5.3 mostren el comportament de les variables duals i com es van adaptant fins assolir el valor òptim.

Finalment, la restricció de potència queda reflectida a la figura 5.4. En ella es veu la potèn-

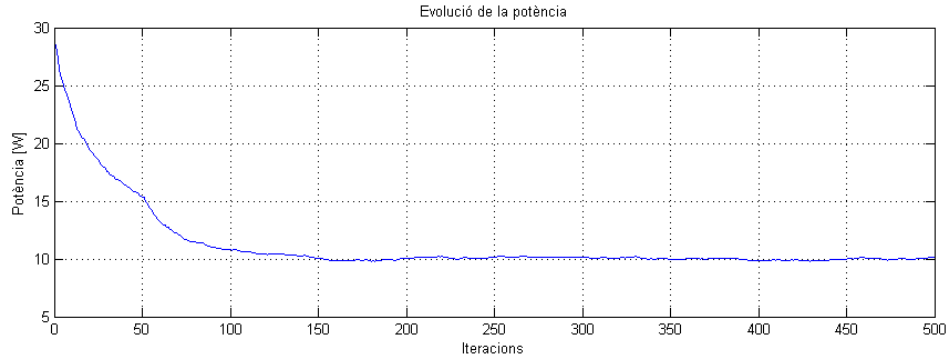


Figura 5.4: Evolució amitjanada de la potència.

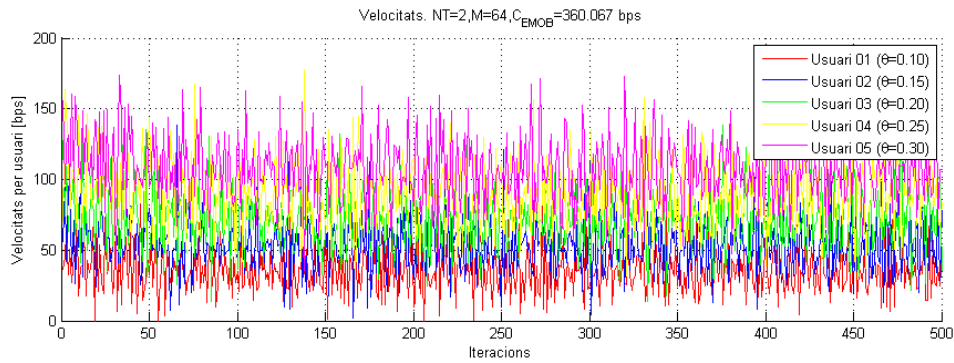


Figura 5.5: Velocitats instantànies per usuari.

cia com va convergint a la potència mitja del sistema. La potència és inversament proporcional a la λ . Per tant, el valor inicial de la potència ve determinat pel valor inicial de la λ .

Cal tenir present que les figures 5.1, 5.3 i 5.4 estan amitjanades, ja que l'objectiu a maximitzar és ergòdic, en valor mig. Però instantàniament varien i sofreixen una variància a tenir en compte. Les velocitats no són constants; les variables μ_k tampoc ho són, per adaptar la velocitat a la restricció a partir del *sumrate*; i la potència primerament sofreix una adaptació inicial fins el valor esperat i seguidament oscil·la al voltant d'aquest mateix. Aquestes oscil·lacions són importants perquè permeten aprofitar les condicions del medi òptimament. És a dir, aprofiten la diversitat d'usuari assignant més potència instantàniament si en un moment donat se'n treu benefici i reduint-la posteriorment quan les condicions són hostils.

Les figures 5.5, 5.6 i 5.7 mostren els resultats instantàniament. Mostren les oscil·lacions esmentades que permeten l'adaptació.

En canvi, la λ presenta els valors instantàniament, sense amitjanar. L'evolució és molt més lenta i no és necessari realitzar una mitjana. Tot i que també oscil·la per seguir les variacions de la potència, és tan influent que amb poca variació ja modifica de forma significant la potència.

Un altre punt a tenir en compte, són les oscil·lacions al voltant del valor mig que permeten a l'algoritme readaptar-se si canvien les circumstàncies. Això inclou l'estadística del canal, les prioritats, l'entrada de nous usuaris, la sortida d'usuaris existents, canvi de prioritats al vol,

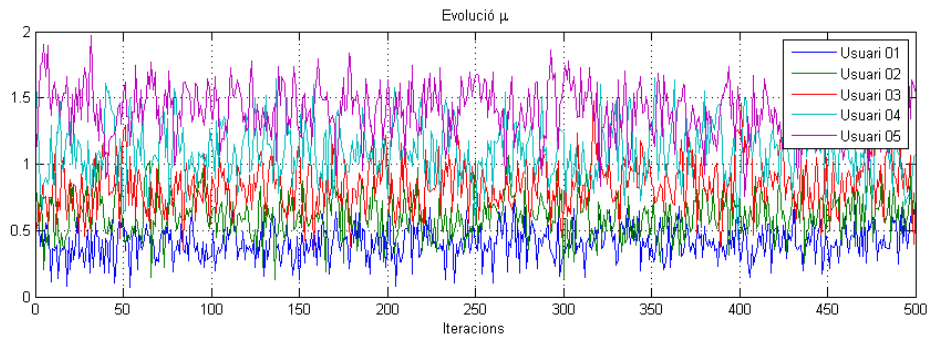


Figura 5.6: Evolució instantània de la μ .

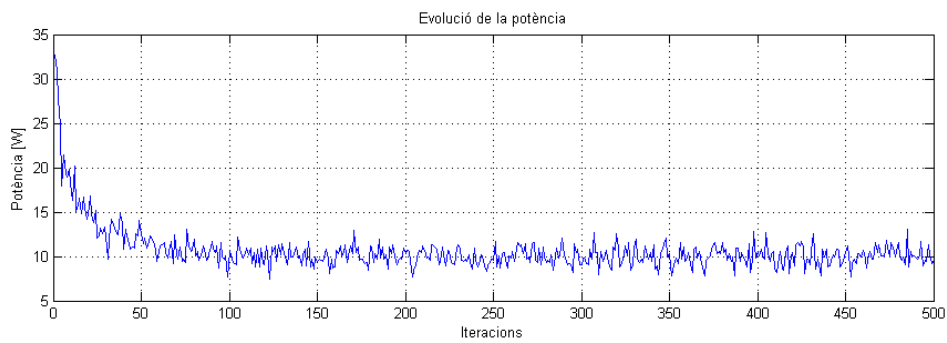


Figura 5.7: Evolució instantània de la potència.

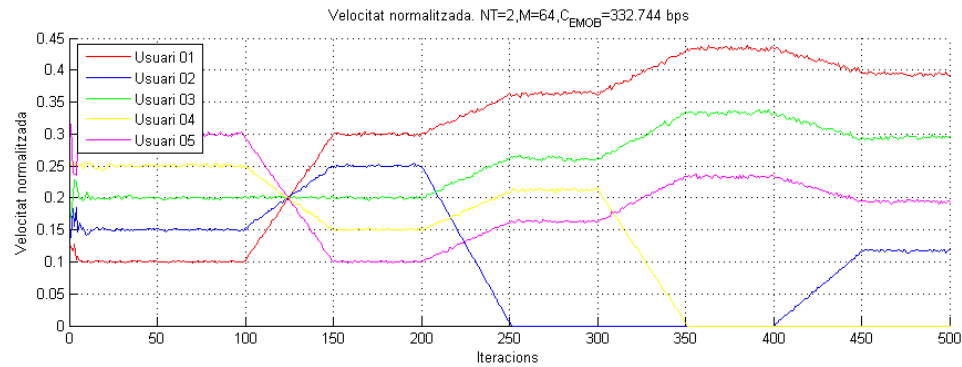
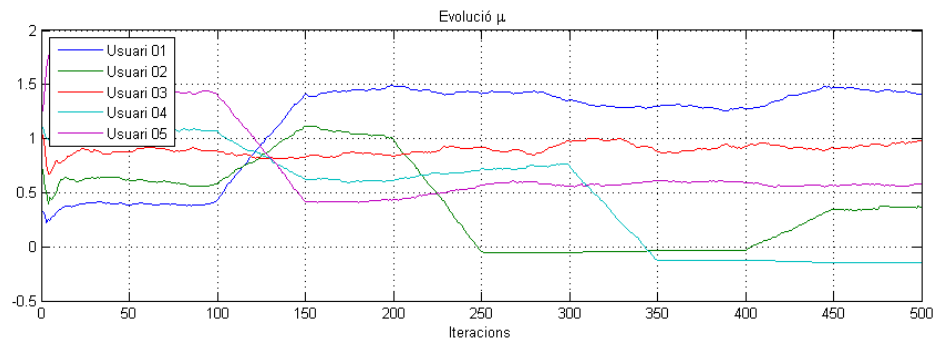
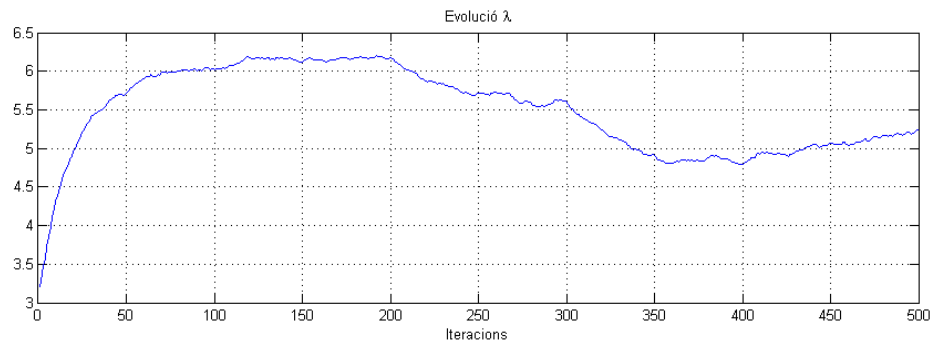


Figura 5.8: Velocitat mitja per usuari amb canvi d'usuaris.

Figura 5.9: Evolució amitjanada de la μ amb canvi d'usuaris.Figura 5.10: Evolució instantània de la λ amb canvi d'usuaris.

etc.

Les figures 5.8, 5.9, 5.10 i 5.11 mostren com el sistema és capaç de readaptar-se quan es produeix un canvi a l'escenari. Concretament s'ha simulat una redistribució de les prioritats entre els usuaris (iteració 100), la sortida de dos usuaris en moments diferents (iteracions 200 i 300) i l'entrada d'un usuari nou (iteració 400).

El sistema és capaç de seguir els canvis i d'adoptar les mesures que requereixi per fer front a la demanda. A la figura 5.8 s'observa com els usuaris guanyen velocitat quan surt un

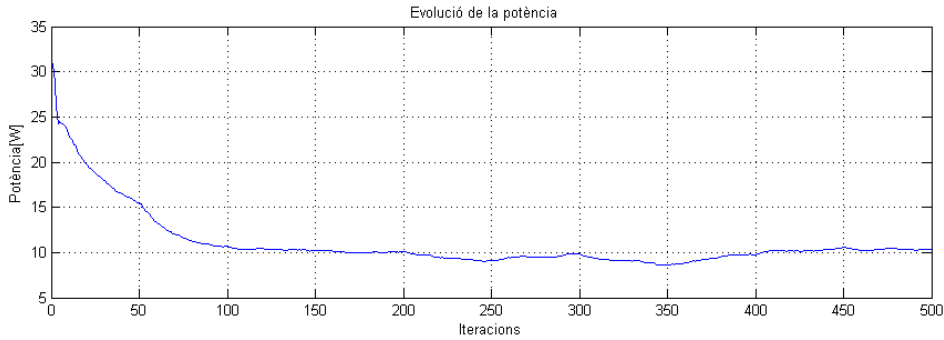


Figura 5.11: Evolució amitjanada de la potència amb canvi d'usuaris.

usuari i en perden quan n'entra un de nou. El règim transitori quan hi ha un canvi quasi és imperceptible. Les figures estan mostrades amb un amitjanat en finestrat de 50 mostres. Per aquest motiu no és fins les 50 mostres que assoleix l'estat estacionari. Les figures amb les velocitats instantànies mostren més fidelment la convergència, encara que no es percep amb claredat el valor mig.

La figura 5.9 reflecteix com es van redistribuint els multiplicadors μ_k en funció de la velocitat i de les prioritats.

La figura 5.10 mostra com el multiplicador λ es va modificant quan entra o surt un usuari. En canvi, roman constant quan es modifica una prioritat. Això resulta obvi perquè la λ només controla la restricció de potència i no es veu afectada quan es modifica el vector de prioritats ϕ . En altres paraules, quan es modifiquen les prioritats el *sumrate* no es veu alterat, ja que és la suma de tots. Però quan entra o surt un usuari el *sumrate* es veu modificat i, per tant, la potència també es veu afectada d'alguna forma o altra.

Finalment, la figura 5.11 ensenya com la potència, un cop s'ha adaptat, no sofreix modificacions del seu valor esperat. Per tant, ni la redistribució de prioritats ni l'entrada ni sortida d'usuaris modifica el valor mig de la potència assignada pel sistema.

En valors instantanis es pot apreciar que ràpidament es produeix l'adaptació i que quasi no hi ha règim transitori. Tot i que les oscil·lacions són pronunciades, a la figura 5.12 s'aprecia com la velocitat s'incrementa quan surt un usuari i decreix quan n'entra un de nou.

De la mateixa manera, la figura 5.13 mostra com els multiplicadors es reajusten ràpidament quan entra o surt un usuari per modificar la velocitat. Noti's com els multiplicadors es fan nuls quan surt un usuari del sistema.

Finalment, la figura 5.14 mostra com la potència instantània quasi no sofreix l'entrada ni sortida d'usuaris i s'adapta ràpidament al valor mig del sistema.

5.1.2 Comparativa

El següent pas és comparar l'algoritme amb altres implementacions més i menys òptimes. Com a referent més òptim s'ha escollit el *Dirty Paper Coding*, donat que aporta la solució més òptima. El DPC ofereix una solució i troba la capacitat del canal MIMO en entorns *broadcasting*. La implementació escollida és [KOB08] perquè permet introduir pesos i prioritats, a part de ser una implementació més o menys senzilla, tot i la complexitat resultant. No obstant, aquesta publicació no contempla la diversitat freqüencial. Per tant, per adaptar-ho a l'escenari plantejat es fa una assignació uniforme de la potència entre les portadores. És a dir,

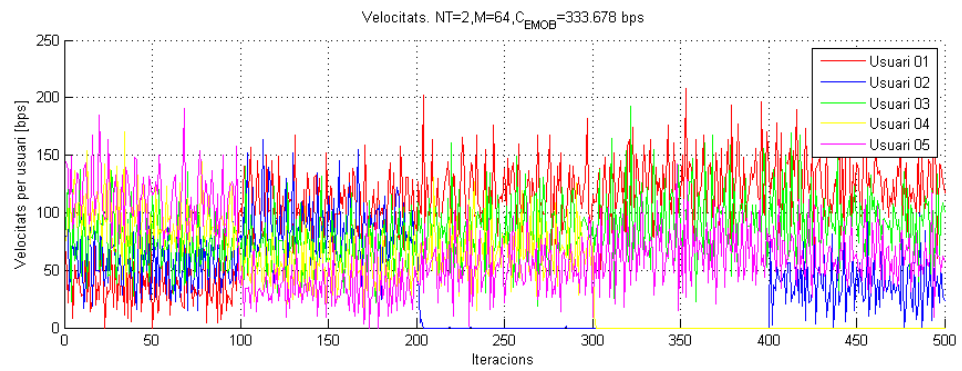


Figura 5.12: Velocitat instantània per usuari amb canvi d'usuaris.

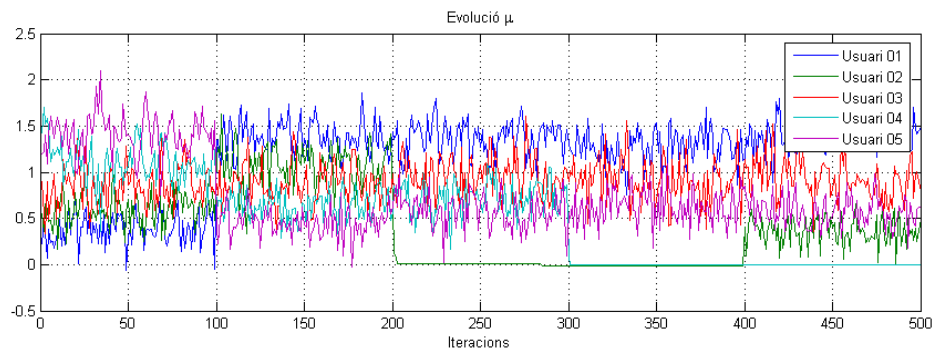


Figura 5.13: Evolució instantània de la μ amb canvi d'usuaris.

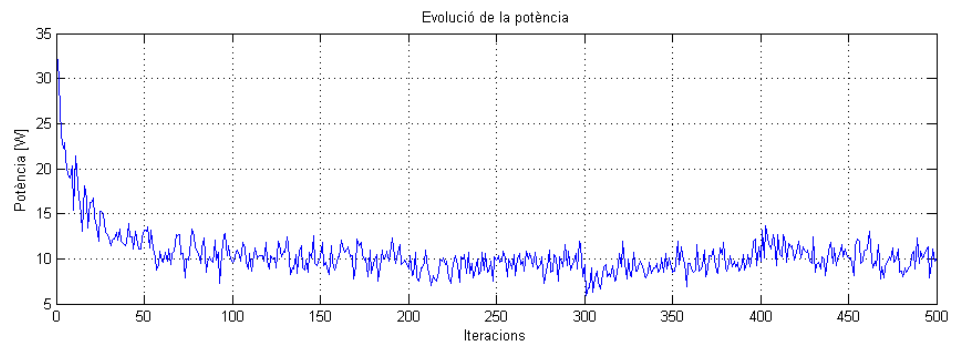


Figura 5.14: Evolució instantània de la potència amb canvi d'usuaris.

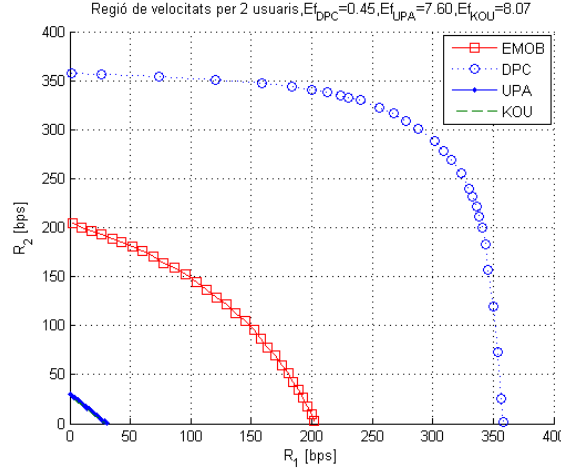


Figura 5.15: Regió de velocitat per dos usuaris amb múltiples implementacions.

que cada portadora se li assigna $P_m = \frac{\bar{P}}{M}$ i s'aplica la solució proposada. Ja es pot apreciar que el resultat que s'obté és en realitat subòptim i es perd eficiència en termes de velocitat. No obstant, serveix per marcar una cota i saber quant lluny s'està d'aquest límit òptim.

Per l'altra banda, com a implementació més subòptima però menys complexa s'ha utilitzat la proposada per [KOU07]. Aquesta implementació resulta molt més senzilla però menys eficient.

La figura 5.15 mostra la implementació de l'algoritme proposat amb les comparatives mencionades, etiquetades com a DPC i KOU respectivament. La implementació UPA correspon a la distribució uniforme de potència entre freqüències i usuaris. UPA és la més senzilla d'implementar ja que no cal realitzar cap recerca.

Al títol superior de la imatge es parla d'eficiència. S'ha definit l'eficiència com

$$Ef_{DPC} = \frac{\bar{C}_{sum}^{EMOB}}{\bar{C}_{sum}^{DPC}} \quad (5.2)$$

$$Ef_{KOU} = \frac{\bar{C}_{sum}^{EMOB}}{\bar{C}_{sum}^{KOU}} \quad (5.3)$$

$$Ef_{UPA} = \frac{\bar{C}_{sum}^{EMOB}}{\bar{C}_{sum}^{UPA}} \quad (5.4)$$

i indica la proximitat amb la implementació comparada. Comparant-lo amb DPC, es mostra com l'eficiència està a la meitat. És a dir, que s'assoleix la meitat del que s'aconseguiria utilitzant DPC. Però òbviament cal tenir present la complexitat. Tot i obtenir la meitat de velocitat, és molt menys complex i requereix molt menys *feedback*.

En canvi, les altres estratègies, tot i ser menys complexes disten força de les velocitats que s'aconsegueixen. Aproximadament, amb una mica menys de complexitat es queden vuit vegades per sota del que s'aconsegueix amb l'algoritme proposat. Es posa de manifest, doncs, que la solució resulta ser atractiva per distar només de la meitat del límit òptim i ser molt

menys complex i, a més, ser vuit vegades més eficient que altres solucions una mica menys complexes.

La complexitat de la solució proposada per [KOB08] té una complexitat força elevada. Concretament, de $\mathcal{O}(K^2 M \log N)$, on $\log N$ és una complexitat logarítmica d'una recerca exhaustiva de N mostres. Com s'aprecia, és molt més complex i requereix molt més temps per computar-ho. Per això, s'utilitza per calcular el màxim teòric que es podria arribar a obtenir.

D'altra banda, la solució formulada per [KOU07] només té una complexitat $\mathcal{O}(M)$. És menys complex però l'actuació és molt pitjor.

Per tant, la solució proposada resulta un intermedi entre les solucions més i menys òptimes i més i menys complexes. S'aconsegueix una bona actuació amb relativament poca complexitat.

D'altra banda, un aspecte important que s'ha remarcat constantment en aquest projecte és la baixa complexitat. El seu anàlisi resulta molt subjectiu perquè, en valor absolut, està subjecta a paràmetres com el rendiment del processador, sistema operatiu, etc. Expressar el consum de temps per realitzar una acció en temps aporta poca informació perquè no mostra en sí mateix si té molta o poca complexitat. La complexitat es mostra a la taula 4.1 però no indica, en unitats de temps, de quant es tracta.

Per tant, oferir resultats expressats en unitats de temps només aporta informació quan es comparen amb altres i es pot apreciar realment els diferents tipus de complexitat lineal, exponencial o quadràtica. La figura 5.16 compara, en segons, el temps que necessita l'algoritme per realitzar la seva funció amb les modalitats de $Q_m = N_T$ o bé Q_m com a variable (etiquetat com *dyn* Q_m), a més de fer-ho per 8 i 32 usuaris, en funció del nombre d'antenes. Com s'ha detallat en el capítol anterior, la complexitat depèn del nombre d'usuaris, de portadores i d'antenes. El que varia, però, és el comportament lineal o exponencial segons la configuració. El valor en segons en sí no és rellevant, ja que està en funció del processador, de la càrrega de la màquina, del sistema operatiu, etc.

El que cal remarcar d'aquesta figura és, en primer lloc, el comportament exponencial quan s'utilitza la configuració dinàmica. S'aprecia amb detall com l'envoltant de les barres dinàmiques segueix una corba exponencial.

Per altre lloc, també cal prestar atenció al comportament lineal entre les barres de la mateixa configuració, ja que mostren la tendència lineal en funció del nombre d'usuaris. I realment és així. Totes les configuracions són lineals amb el nombre d'usuaris.

Així doncs, encara que la magnitud del temps és irrellevant, reflecteix clarament les corbes de complexitat expressades a la taula 4.1.

El nombre d'antenes és un paràmetre que porta amb sí mateix una gran importància. Defineix la complexitat del sistema i, a més, influeix directament en l'optimització. Augmentar-lo no significa necessàriament que augmenti el *sumrate*. Això és perquè si s'augmenta el nombre d'antenes també s'augmentarà el nombre d'usuaris a multiplexar i, per tant, la interferència. Òbviament aquesta apreciació està subjecta a l'estratègia que s'utilitza. Deixant el nombre d'usuaris a servir espacialment (Q_m) com una variable, optimitzarà el resultat i, llavors, augmentant el nombre d'antenes sí farà créixer el *sumrate*.

Això es mostra a la figura 5.17. El que primer s'aprecia és la contribució de la configuració en el *sumrate* en funció del nombre d'antenes. S'observa clarament com, efectivament, deixant el paràmetre d'usuaris a servir simultàniament o nombre de *beams*, Q_m , com a variable s'aconsegueix augmentar el *sumrate* proporcionalment al nombre d'antenes. Això cobra sentit perquè d'aquesta manera es busca aquella combinació que maximitza la SINR i, per tant, fa augmentar la velocitat total del sistema. Per contra, si aquest paràmetre es deixa fix s'obté un

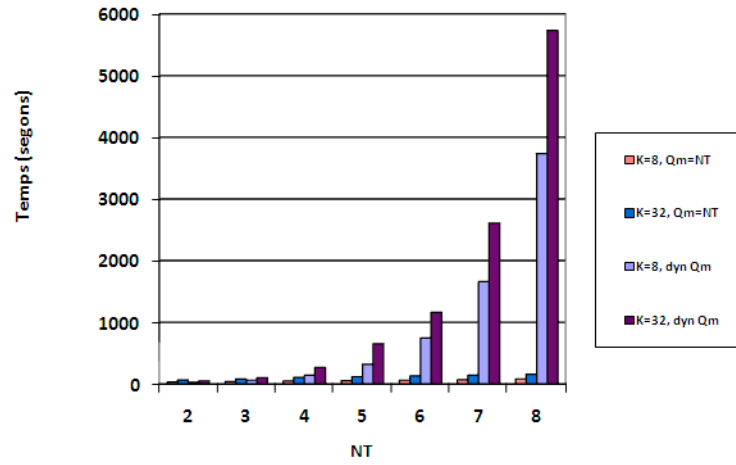


Figura 5.16: Consum de temps de l'algoritme en diferents configuracions i usuaris.

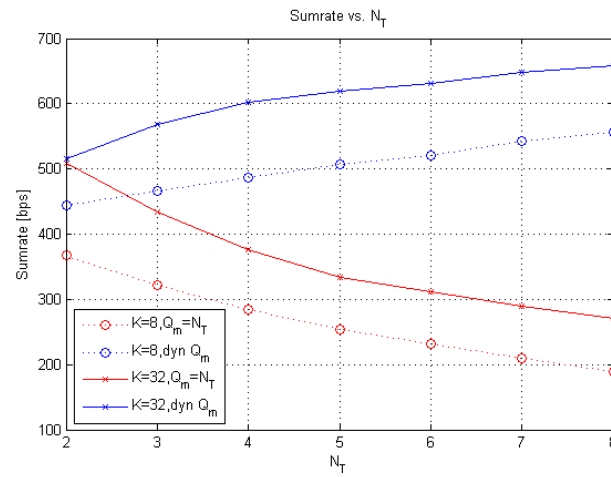


Figura 5.17: *Sumrate* en funció del nombre d'antenes per $K = 8$ i $K = 32$ usuaris i múltiples configuracions.

empobriment del *sumrate*. Amb el paràmetre fix, quan s'augmenta el nombre d'antenes també augmenta la interferència innegablement i fa disminuir la SINR, provocant un decreixement de la velocitat global.

Per tant, augmentar el nombre d'antenes no sempre és una bona solució. L'avantatge que aporta fer-ho és la possibilitat de servir a més usuaris simultàniament obtenint, però, menys velocitat global. No obstant, si, a més, també es vol augmentar el *sumrate*, llavors s'ha de deixar el paràmetre lliure, això sí, guanyant complexitat.

Com es veu, hi ha múltiples configuracions en funció dels propòsits que es busquin: augmentar el *sumrate*, servir un gran nombre d'usuaris simultàniament o reduir la complexitat. Evidentment totes tres premisses no es poden obtenir i es crea un compromís.

Un altre factor important que es desprèn de la figura 5.17 és el *sumrate* en funció del nombre d'usuaris. En aquest cas, augmenta si també ho fa el nombre d'usuaris. Això es desprèn del concepte de la conformació oportunista per explotar la diversitat d'usuari. Augmentant el nombre d'usuaris s'augmenten les possibilitats de trobar un millor canal i, per tant, esdevenir un *sumrate* molt major, tal i com succeeix. El guany que s'obté és la diferència entre les corbes del mateix color. I, a més a més, aquest guany resulta pràcticament el mateix, sigui la configuració que sigui. Això és perquè aquest guany és independent de la configuració adoptada i es deriva d'utilitzar l'estratègia oportunista.

Finalment, un altre aspecte que es desprèn és el comportament. Les corbes segueixen funcions logarítmiques i això fa pensar que la influència important del nombre d'antenes es produeix amb un nombre baix. Amb dues o tres antenes ja s'aprecien resultats notables. En canvi, per l'altre extrem, es necessita un nombre elevat per fer variar substancialment el *sumrate*.

5.2 Escenari real

Les següents simulacions estan realitzades en un escenari més real¹. Aquest escenari incorpora aspectes més reals que s'adeqüen millor a situacions pràctiques de les noves comunicacions. Sistemes com Wifi [WIF30], WiMAX [WIM29] o 3GPP-LTE utilitzen la tecnologia OFDMA com a mecanisme de permís d'accés.

Es considera una sola cèl·lula, sense interferència intercel·lular. Els paràmetres d'OFDMA són $M = 128$ portadores, amb un ample de banda de 1.25MHz . El sistema té un multiplexatge per temps, TDD, on $2/5$ del temps d'interval s'utilitza per l'enllaç descendent. Els paràmetres de CSI s'actualitzen cada 10ms . Això significa que entre els intervals de n transcorren 10ms . En aquest temps, l'estació base sondeja els diferents canals, els usuaris reenvien el *feedback*, es realitza la gestió d'accés i la transmissió de potència. Es contempen dues situacions diferents de la distribució d'usuaris. En la primera, els usuaris es distribueixen uniformement en una àrea circular de 0.78km^2 o 500m de radi. En la segona, els usuaris tots es situen a una distància de 250m de l'estació base. El model de canal inclou pèrdues d'atenuació per distància o *path-loss*, esvaïments ràpids i esvaïments lents (no presents a la segona situació). El *path-loss* es modela segons la funció $L(\text{dB}) = 40 + 15.1 \log(d)$, essent d la distància a l'estació base. Els esvaïments lents tenen una desviació típica de $\sigma = 6\text{dB}$. Els esvaïments ràpids segueixen l'estadística de Rayleigh. També s'inclou una correlació deguda al moviment, d'un ample de banda de Doppler de 6Hz . Finalment, existeix un nombre finit de múltiples camins per on arriba el senyal, amb un retard mig de $2 - 3\mu\text{s}$ i de potència normalitzada a la unitat. L'estació

¹Les simulacions estan realitzades a ENDIF - CNIT d'Itàlia sota la supervisió del Dr. Velio Tralli.

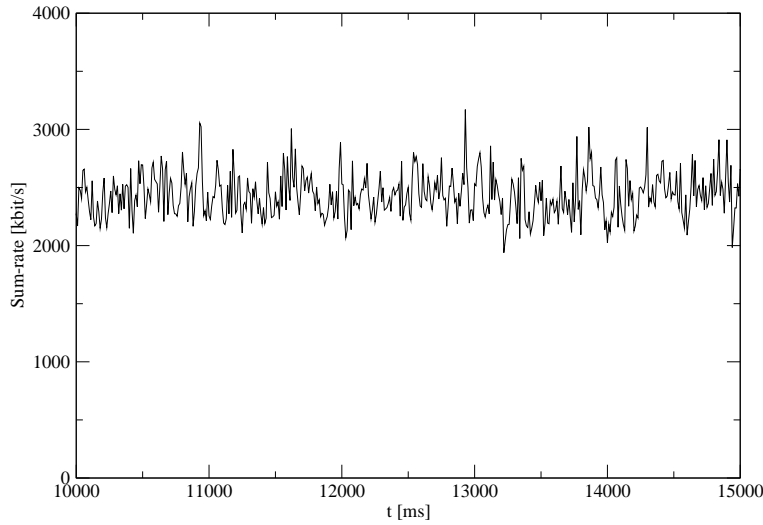


Figura 5.18: *Sumrate* per 10 usuaris a 250m de distància.

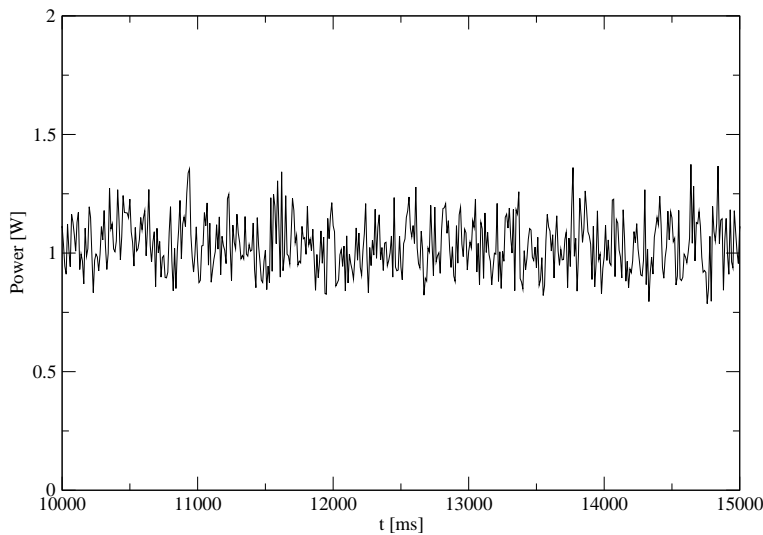


Figura 5.19: Evolució de la potència total per 10 usuaris a 250m de distància.

base genera *beams* ortonormals en cada temps d'interval. Per les simulacions, la potència mitja s'ha fixat a $\bar{P} = 0dB$.

La figura 5.18 mostra el *sumrate* instantani en un sistema real amb l'algoritme incorporat. Com s'aprecia, oscil·la al voltant d'un valor mig, en funció de les condicions estacionàries de l'escenari. Al mateix temps, la figura 5.20 també reflecteix el *sumrate* instantani però per 20 usuaris. S'observa com s'incrementa degut a la diversitat multiusuari.

Finalment, la figura 5.19 mostra com la potència queda adaptada de seguida al valor fixat.

Aquestes simulacions són útils per referendar els resultats anteriors en sistemes i estàndards ja implementats. Tal com s'ha vist, les simulacions concorden amb el que s'espera obtenir.

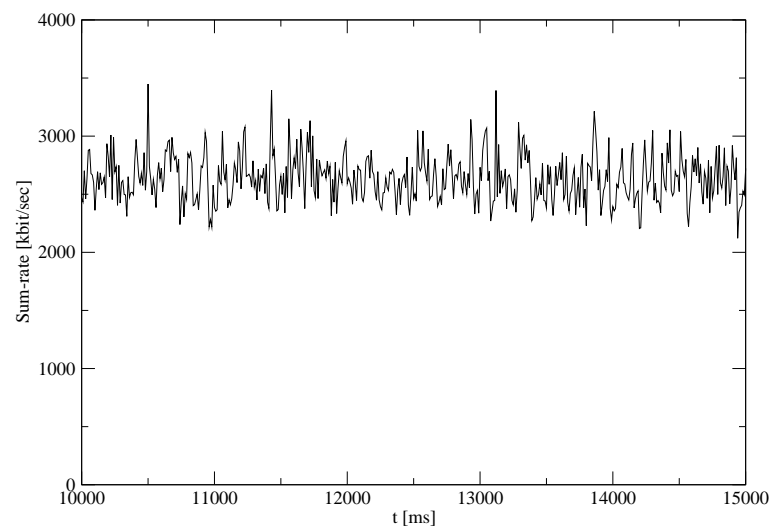


Figura 5.20: *Sumrate* per 20 usuaris a 250m de distància.



6 Conclusions i treball futur

Els sistemes multiusuari certament presenten una gran complexitat. Per sí mateixos ja es fa difícil trobar una cota o un màxim que es pot assolir. I en aquests contextos, la complexitat computacional és molt elevada. Trobar alternatives que satisfacin el compromís entre optimització i complexitat és una tasca complicada.

Aquest projecte ha pretès mostrar una via de poca complexitat de càlcul i que, a més, sigui quasi òptima. Evidentment mai s'ha volgut arribar al màxim que pot presentar el DPC però sí aproximar-s'hi mantenint sempre la simplicitat. Les bones idees són aquelles que, a més de ser eficients, són senzilles. I és el que aquest projecte sempre ha volgut mantenir.

S'han presentat els fonaments necessaris perquè qualsevol lector no familiaritzat amb la temàtica pugui arribar a entendre tot el projecte. També s'ha introduït la problemàtica que presenten aquests tipus d'escenaris. I, seguidament, s'ha mostrat amb detall la recerca que s'ha dut a terme per fer front a aquesta demanda. Els resultats obtinguts encoratgen a seguir investigant noves possibilitats i nous escenaris. A més, la bona acollida que han tingut en els diferents congressos on s'han publicat fa que també hi hagi un bon optimisme al respecte.

A mode de conclusions es volen destacar, primordialment, la quasi optimització i la proximitat respecte el DPC i la poca complexitat. Aquests dos punts s'han de mostrar conjuntament, ja que no es pot parlar de grau d'optimització sense indicar la complexitat que requereix. A més, també es destaca la incorporació de prioritats. Aquest fet fa que s'obri la possibilitat d'incorporar QoS al sistema i facilitar el tràfic heterogeni, tal i com mostra la publicació despresa [TRA18].

El fet d'unir la diversitat freqüencial que ofereix OFDMA amb la diversitat espacial que presenta l'esquema oportunista i la diversitat temporal que incorpora la maximització ergòdica, desencadena en aquest algoritme per oferir un bon resultat a baix cost. Les dues primeres ofereixen més prestacions en termes de velocitat i l'última permet relaxar la complexitat.

Aquests aspectes fan que es plantegin altres interrogants i àrees que resulten atractives d'investigar. La primera d'elles és estudiar l'opció que els terminals mòbils incloguin també múltiples antenes. En aquests cas, desapareix el concepte de SINR o CINR instantània que percep un sol usuari. Per tant, cal indagar una manera d'abordar aquest problema i que es pugui integrar amb el que s'ha plantejat.

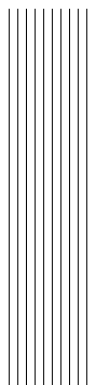
Un altre tema que es mostra interessant és estudiar amb més profunditat la implementació de QoS amb tràfic heterogeni. En aquest cas, la demanda i accés del medi és molt més variable i molt més diversa que en els escenaris estudiats. Tot i que els resultats són esperançadors perquè indiquen que el sistema s'adapta ràpidament cal realitzar un estudi més exhaustiu.

També es pot incorporar un estudi que centri els seus esforços en realitzar el mateix raonament però amb *rate* discret. Això equival a plantejar el problema inicial d'aquest projecte però també afegint la variable del tipus de modulació. Realitzant una modulació adaptativa es podrà aconseguir uns valors discrets de la velocitat. No obstant, cal estudiar-ho amb deteniment i trobar-hi la manera d'incorporar-ho.

En tot el projecte s'ha treballat sota la condició de conèixer el CSI parcial tant en transmissió com en recepció. Això ajudaria a valorar l'impacte que tindria un CSI imperfecte. És a dir, estudiar com afecten els errors de mesura i d'imprecisió que es cometen durant l'estimació del canal al receptor.

Finalment, una altra opció que es desprèn de la idea principal és la de realitzar estudis amb la capacitat d'*outage* i minimització de la potència ergòdica. Aquest plantejament difereix del problema inicial però també resulta interessant perquè permetria relaxar les restriccions.

Com hom pot apreciar, el ventall que es desplega és important i ofereix moltes noves opcions que reclamen ser estudiades amb deteniment. Tot i que el que s'ha presentat és en sí mateix el nucli d'aquest tipus de marc de treball, cal tenir-lo present en les posteriors investigacions que es duguin a terme.



Apèndixs



A Apèndix A. Apèndix matemàtic

En aquest apèndix es mostren les apreciacions matemàtiques que requereixen una especial atenció i no són incloses en els capítols perquè desviarien l'atenció del tema central.

A.1 Teoria matricial

Aquest projecte fa un fort ús de les matrius i vectors. Per aquesta raó es detallen a continuació alguns aspectes importants relatius al càlcul matricial.

A.1.1 Autovalors i autovectors

Sigui la matriu $\mathbf{A} \in \mathbb{C}^{n \times n}$, el vector $\mathbf{v} \in \mathbb{C}^n$ i l'escalar λ que satisfan l'equació

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}. \quad (\text{A.1})$$

Si la solució és tal que $\lambda \neq 0$ i $\mathbf{v} \neq \mathbf{0}$, llavors λ s'anomena autovalor de \mathbf{A} i \mathbf{v} , autovector de \mathbf{A} associat a λ .

Dit d'altra manera, λ és un autovalor de \mathbf{A} si i només si

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0. \quad (\text{A.2})$$

El vector associat \mathbf{v} es pot trobar resolent l'equació

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}. \quad (\text{A.3})$$

A.1.2 Descomposició en valors singulars. SVD

La descomposició SVD està associada a la factorització en autovalors i autovectors d'una matriu hermitica¹. Donat que els autovectors d'una matriu hermitica són ortonormals entre sí, es pot reescriure \mathbf{A} com

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^H \quad (\text{A.4})$$

¹Una matriu \mathbf{A} és hermitica si $\mathbf{A} = \mathbf{A}^H$.

on \mathbf{D} és la matriu diagonal amb els autovalors. A més, \mathbf{U} és una matriu unitària, és a dir, $\mathbf{U}^H \mathbf{U} = \mathbf{I}$.

No obstant, la descomposició SVD es pot aplicar a qualsevol matriu. Sigui la matriu $\mathbf{A} \in \mathbb{C}^{n \times m}$. La seva factorització resulta

$$\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{V}^H. \quad (\text{A.5})$$

Les columnes de la matriu $\mathbf{U} \in \mathbb{C}^{n \times n}$ són els autovectors de $\mathbf{A} \mathbf{A}^H$ i les columnes de \mathbf{V} són els autovectors de $\mathbf{A}^H \mathbf{A}$. Els valors singulars es situen a la diagonal de la matriu $\mathbf{D} \in \mathbb{C}^{n \times m}$ i són les arrels quadrades dels autovalors de $\mathbf{A} \mathbf{A}^H$ i de $\mathbf{A}^H \mathbf{A}$.

Es pot demostrar a partir de les igualtats

$$\begin{aligned} \mathbf{A} \mathbf{A}^H &= \mathbf{U} \mathbf{D} \mathbf{V}^H (\mathbf{U} \mathbf{D} \mathbf{V}^H)^H \\ &= \mathbf{U} \mathbf{D} \mathbf{V}^H \mathbf{V} \mathbf{D}^H \mathbf{U}^H \\ &= \mathbf{U} \mathbf{D} \mathbf{D}^H \mathbf{U}^H \\ &= \mathbf{U} \mathbf{D}^2 \mathbf{U}^H \end{aligned} \quad (\text{A.6})$$

De la mateixa manera, es pot demostrar per la matriu \mathbf{V} .

Es defineix el rang d'una matriu com el nombre d'autovalors diferents de zero.

A.2 Propietats emprades

1. Matriu hermítica: $\mathbf{A} = \mathbf{A}^H$.
2. $\|\mathbf{v}\|^2 = \mathbf{v}^H \mathbf{v} = \text{tr}(\mathbf{v} \mathbf{v}^H)$.
3. $\det(\mathbf{I} + \mathbf{A} \mathbf{B}) = \det(\mathbf{I} + \mathbf{B} \mathbf{A})$.
4. $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$.
5. $\text{tr}(\mathbf{A} \mathbf{B}) = \text{tr}(\mathbf{B} \mathbf{A})$.

A.3 Derivada de la funció objectiu

La solució de l'equació 4.6 es troba a partir del conjunt de $\{p_{k,m}\}$ que maximitza la funció objectiu. Per un $p_{k,m}$ arbitrari, maximitza la funció si

$$\frac{\partial \Theta(\lambda, \boldsymbol{\mu})}{\partial p_{k,m}} = 0 \quad (\text{A.7})$$

que és el mateix que

$$\frac{\partial \Theta_k(\lambda, \boldsymbol{\mu})}{\partial p_{k,m}} = 0. \quad (\text{A.8})$$

Tenint en compte que la potència instantània o en la iteració i no fa ús de l'esperança, llavors es té que

$$\frac{\partial \Theta_k(\lambda, \boldsymbol{\mu})}{\partial p_{k,m}} = \mu_k \frac{\xi_{k,m}}{1 + \xi_{k,m} p_{k,m}} \cdot \frac{1}{\ln 2} - \lambda = 0. \quad (\text{A.9})$$

Manipulant l'expressió anterior es pot escriure

$$\frac{\mu_k}{\lambda \ln 2} - \frac{1}{\xi_{k,m}} = p_{k,m} \quad (\text{A.10})$$

que concorda amb la solució òptima.

Estenent l'expressió anterior per tota la matriu de potències \mathbf{P} , i definint l'operació $\mathbf{B} = \frac{1}{\mathbf{A}}$ com la matriu on cada element $b_{ij} = \frac{1}{a_{ij}}$, llavors la solució òptima resulta

$$\mathbf{P} = \frac{1}{\lambda \ln 2} (\mathbf{1}_M \otimes \boldsymbol{\mu}^H) - \frac{1}{\boldsymbol{\Xi}} \quad (\text{A.11})$$

on $\boldsymbol{\Xi}$ és la matriu de CINR.

A.4 Funcions de probabilitat

En el segon capítol s'ha introduït el concepte d'esvaïment. Aquest fenomen succeeix per múltiples factors com la propagació multicamí, el moviment del terminal o l'aparició d'objectes de grans dimensions. Aquests successos ocorren de manera aleatòria però segueixen una determinada distribució de probabilitat.

A continuació s'examina amb més detall els diferents tipus més habituals de distribució que es poden trobar en escenaris reals.

A.4.1 Rayleigh

Tal i com s'ha pogut apreciar, els esvaïments ràpids són causats principalment per la propagació multicamí. L'amplitud d'aquests coeficients dels múltiples camins segueixen una estadística anomenada Rayleigh.

La funció de densitat de probabilitat (fdp) és

$$p(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}, \quad r \geq 0 \quad (\text{A.12})$$

on $r = |r(t)|$ és l'amplitud de la senyal rebuda i $\sigma^2 = \frac{\mathbb{E}\{r^2(t)\}}{2}$. Aquesta funció mostra la probabilitat de que l'amplitud sigui una o altra. Es pot apreciar amb més detall a la figura [A.1](#).

A.4.2 Rice

La distribució Rice consisteix en una generalització de l'anterior. Aquesta es produeix quan el terminal mòbil està pròxim a l'estació base. Això fa que sigui molt probable que existeixi una línia de visió directa entre el terminal i l'estació. Això provoca que hi hagi un camí que es rebit amb molta més potència que els altres.

D'aquesta manera, es pot dir que un canal Rice conté una component de LOS i una altra component Rayleigh. Això equival a formular que la component LOS s'obté a partir dels vectors de *steering* més el canal Rayleigh habitual. Es pot sintetitzar a partir de

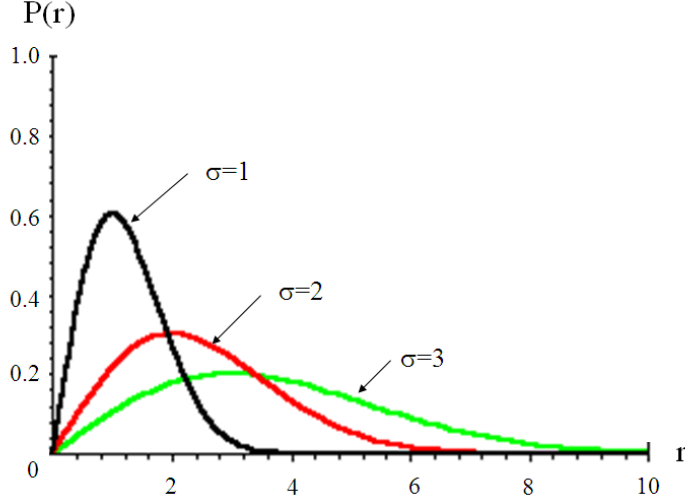


Figura A.1: Funcions de densitat de probabilitat Rayleigh per diferents variàncies.

$$h_{Rice} = h_{LOS} + h_{Rayleigh} = h_{LOS} + h_{Gauss} + j \cdot h_{Gauss}. \quad (A.13)$$

La fdp d'aquesta distribució és

$$p(r) = \frac{r}{\sigma^2} e^{-\frac{r^2 + \alpha^2}{2\sigma^2}} I_0\left(\frac{r\alpha}{\sigma}\right) \quad (A.14)$$

on α és l'amplitud del senyal que arriba directament. $I_0(x)$ és la funció de Bessel del primer tipus d'ordre 0. Aquesta és

$$I_\alpha(x) = \sum_{i=0}^{+\infty} \frac{1}{i! \Gamma(i + \alpha + 1)} \left(\frac{x}{2}\right)^{2i + \alpha} \quad (A.15)$$

on $\Gamma(z)$ és la funció Gamma d'Euler i és

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt. \quad (A.16)$$

De manera particular es té

$$I_0(x) = 1 + \frac{x^2}{2^2} + \frac{x^4}{2^2 4^2} + \frac{x^6}{2^2 4^2 6^2} + \dots \quad (A.17)$$

La figura A.2 il·lustra la fdp d'aquesta distribució.

A.4.3 Lognormal

Els esvaïments lents són causats pel moviment del terminal o l'aparició d'objectes de grans dimensions. Quan això succeeix l'amplitud del senyal rebut decau. L'estadística que descriu segueix una distribució lognormal².

²Els paràmetres han d'estar mesurats en W (Watts). Si les unitats són en decibels, llavors segueix una distribució normal.

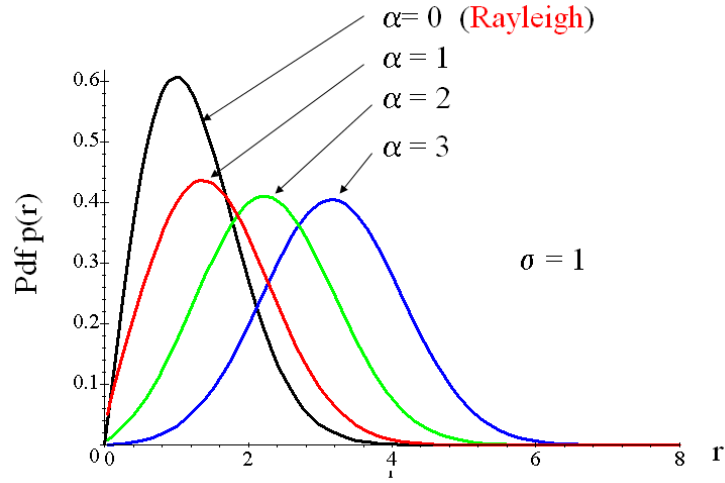


Figura A.2: Funcions de densitat de probabilitat Rice per diferents valors de α .

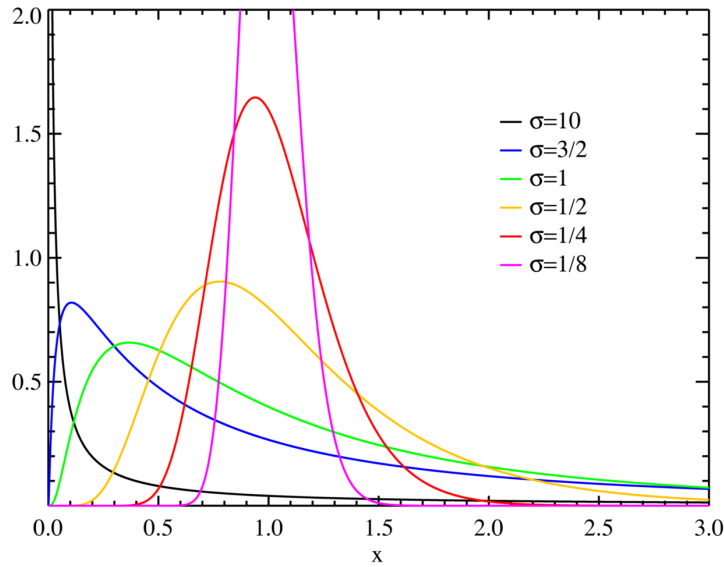


Figura A.3: Funcions de distribució de probabilitat lognormal per diferents variàncies.

La fdp de la distribució lognormal és

$$p(r) = \frac{1}{r\sigma\sqrt{2\pi}} e^{-\frac{(\ln(r)-\bar{r})^2}{2\sigma^2}}. \quad (\text{A.18})$$

La figura A.3 il·lustra el comportament d'aquesta distribució.

Aquest concepte és molt útil per traçar límits o zones de cobertura per assegurar que un determinat percentatge d'ubicacions reben un mínim de potència. Suposant que els valors s'expressen en dB, llavors la probabilitat d'assegurar un mínim de potència ve donada per

$$p(P > P_{th}) = \int_{P_{th}}^{+\infty} f(P) dP = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left(\frac{P_{th} - \bar{P}}{\sqrt{2}\sigma} \right) \quad (\text{A.19})$$

on P_{th} és el mínim de potència que es vol rebre. Es recorda que la funció d'error $\operatorname{erf}(x)$ és

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \quad (\text{A.20})$$

A.5 Optimització dual

L'optimització dual constitueix una eina primordial a l'hora de resoldre problemes d'optimització. La idea principal és la de plantejar un problema dual però de menor complexitat. La solució que es troba pel problema dual també constitueix una solució pel problema inicial, tot i ésser diferents. És a dir, es busca un problema paral·lel, amb unes altres variables, tal que tingui els extrems als mateixos punts que el problema inicial.

Normalment els problemes que es volen optimitzar consten d'una funció a maximitzar i un seguit de restriccions. El que es mostra en tot aquest apartat és una particularització d'aquest tipus de problemes aplicats a sistemes OFDM.

Generalment, el problema que es planteja és

$$\begin{aligned} \max \quad & \sum_{m=1}^M f_m(x_m) \\ \text{s.t.} \quad & \sum_{m=1}^M h_m(x_m) \leq \mathbf{\Lambda} \end{aligned} \quad (\text{A.21})$$

on $f_m(\cdot)$ és una funció que no té per què ser necessàriament còncava i $h_m(\cdot)$ és una funció restricció que no té per què ser necessàriament convexa. $\mathbf{\Lambda}$ és el vector de restriccions.

La manera de resoldre aquests problemes és aplicant els multiplicadors. Aquests incorporen les restriccions a l'expressió inicial i així deixen de ser restriccions com a tals. El problema, per tant, queda més relaxat donat que desapareixen les restriccions que sí té el problema inicial. Així doncs, el lagrangià és

$$L(x_m, \boldsymbol{\lambda}) = \sum_{m=1}^M f_m(x_m) + \boldsymbol{\lambda}^T \left(\mathbf{\Lambda} - \sum_{m=1}^M h_m(x_m) \right) \quad (\text{A.22})$$

on $\boldsymbol{\lambda}$ és multiplicador de Lagrange. En aquest punt, es pot observar que el lagrangià descomposa el problema en M subproblemes que fan molt més fàcil trobar una solució pel problema inicial. Així, la funció objectiu dual té la solució

$$g(\boldsymbol{\lambda}) = \max_{x_m} L(x_m, \boldsymbol{\lambda}). \quad (\text{A.23})$$

El problema dual és, per tant,

$$\begin{aligned} \max \quad & g(\boldsymbol{\lambda}) \\ \text{s.t.} \quad & \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned} \quad (\text{A.24})$$

Si $f_m(x_m)$ i $h_m(x_m)$ són còncaves i convexes, respectivament, l'optimització dual garanteix que ambdós problemes, l'inicial i el dual, comparteixen la mateixa solució. Si no és el cas, llavors el problema dual aporta un límit o una cota superior pel problema inicial. La diferència entre el màxim del problema inicial i el mínim del problema dual s'anomena error dual o *duality gap*.

Els teoremes que es desprenen de [WYU31] i [BGI32] permeten assegurar que si el sistema és *time-sharing* llavors el *duality gap* és nul. Això, a la pràctica, no succeeix. En els sistemes OFDM utilitzen un cert nombre de portadores amb un ample concret. Tot i ser pla en tot l'ample de la portadora, es cometten errors en la medició que desencadenen en que el *duality gap* no sigui nul. A més, el fet de “mostrejar” el canal freqüencialment també hi contribueix.

Generalment, si $M \rightarrow \infty$ llavors és equivalent al *time-sharing*. Això implica infinites portadores d'un ample de banda infinitesimal, fet impossible a la pràctica. Un altre fet que es deriva que queda plasmat a [BGI32] és que llavors la funció d'acumulació del canal és contínua, no fa salts.

Per tant, tot i no ser nul, l'error comès és despreciable i es pot afirmar que el *duality gap* és pràcticament nul. Això es pot demostrar de la següent manera. Un fenomen es diu que és *time-sharing* si compleix la propietat:

Propietat 1 *Siguin x_m i y_m dues solucions òptimes del problema amb les restriccions Λ_x i Λ_y , respectivament. Llavors per un $0 \leq \alpha \leq 1$, existeix un conjunt z_m tal que $\sum_m h(z_m) \leq \alpha \Lambda_x + (1 - \alpha) \Lambda_y$, i $\sum_m f_m(z_m) \geq \alpha \sum_m f_m(x_m) + (1 - \alpha) \sum_m f_m(y_m)$.*

OFDM clarament compleix aquesta propietat perquè les portadores es poden assignar un α del temps a x_m i l'altre $(1 - \alpha)$ del temps a y_m . Aquest fet és propiciat perquè els canals adjacents tenen una resposta plana i equivalent. Per tant, només cal demostrar que el *time-sharing* implica que el *duality gap* sigui nul.

Teorema 1 *Si un problema d'optimització dual satisfà la propietat de time-sharing, llavors el duality gap és nul.*

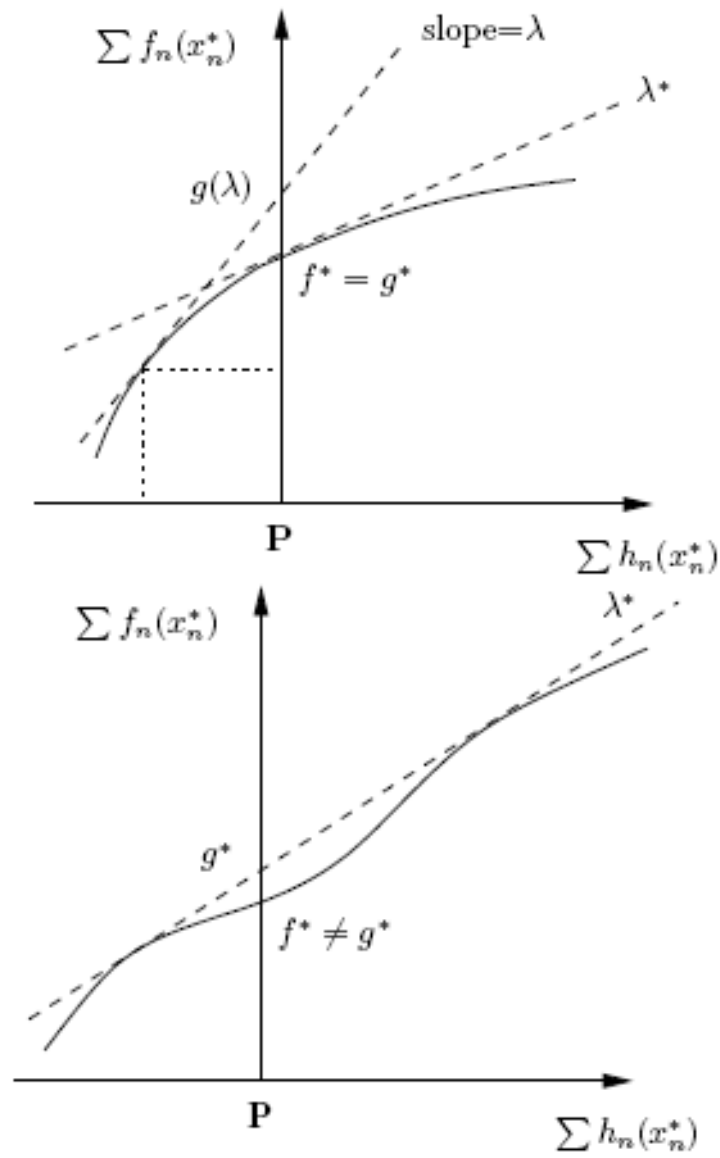
La demostració d'aquest teorema es prova a [WYU31] i a continuació es detalla.

Demostració. El primer diagrama de la figura A.4 mostra aquest comportament. La corba sòlida descriu la trajectòria dels punts $(\sum h_m(x_m^*), \sum f_m(x_m^*))$ per tots els valors òptims de x_m^* . Es pot observar com la intersecció entre la corba i l'eix vertical constitueix la solució òptima ja que en aquest punt $\sum h_m(x_m^*) = \Lambda$. Per tant, la solució al problema dual és la λ que es correspon amb el pendent de la recta tangent (que és $g(\lambda)$) en aquest punt. Com es veu, en aquest punt el *duality gap* és nul perquè tant $g(\lambda)$ com el problema inicial coincideixen.

Un fet a remarcar és que també s'aprecia com la corba és còncava. Si es fa variar λ llavors $g(\lambda)$, la recta tangent, assoleix un mínim exactament amb el punt màxim del problema inicial.

El segon diagrama mostra un problema que no compleix la propietat de *time-sharing*. S'observa com en el punt $\sum h_m(x_m^*) = \Lambda$ la diferència entre la recta tangent $g(\lambda)$ no és nul·la. \square

D'aquesta manera es pot demostrar com la solució dual realment constitueix la solució al problema inicial sempre que es compleixi la propietat anterior. La manera de trobar el valor d'aquest pendent λ es pot realitzar mitjançant el subgradient, com s'ha descrit en els capítols del projecte.

Figura A.4: Conseqüència del *time-sharing*.

A.6 Ergodicitat

Durant tot el projecte es parla de la maximització ergòdica i es suposa per apresada. No obstant, en aquest apartat es farà una breu explicació matemàtica.

Una variable aleatòria pren valors no determinats però que segueixen un ordre, una distribució. Per tant, es pot traçar un valor intermedi sobre el que la variable oscil·la. Aquest valor mig es calcula ponderant tots els valors per la seva probabilitat. És a dir, calculant la seva esperança matemàtica. Sigui χ una variable aleatòria i $f_\chi(x)$ la seva funció de densitat, el seu valor mig és l'esperança matemàtica i s'expressa amb

$$\mu_\chi = \mathbb{E}\{\chi\} = \int_{-\infty}^{+\infty} x f_\chi(x) dx. \quad (\text{A.25})$$

Per conèixer el seu valor mig, per tant, és necessari conèixer a priori la funció de densitat. És a dir, cal saber quin comportament segueix la variable aleatòria. Això requereix una informació addicional que no sempre es té. Pel cas d'aquest projecte, no sempre es té un comportament exacte de l'estadística del canal. Per això, s'incorpora la propietat d'ergodicitat.

Un procés estocàstic és una funció de variables deterministes i de variables aleatòries. Per aquest projecte, els processos estocàstics són en funcions del temps i de la variable aleatòria del canal i del soroll. De la mateixa manera, es pot definir el valor mig del procés a partir de la seva esperança matemàtica.

Propietat 2 *Un procés estocàstic és ergòdic si els seus moments estadístics es poden obtenir mitjançant el processat en el temps d'una única realització del mateix. Concretament, un procés estocàstic és ergòdic en mitjana si aquesta es pot obtenir processant-ne en el temps una realització. És a dir,*

$$\mathbb{E}\{f(t; \chi)\} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} f(t; \chi) dt \quad (\text{A.26})$$

per una realització de χ .

Això equival a formular que el promedi temporal és el mateix que el de la variable aleatòria. Aquest projecte fa ús d'aquesta propietat per calcular la seva esperança matemàtica. Alhora d'implementar l'algoritme, es van obtenint mostres del canal i els resultats es van amitjanant a partir de la mostra actual i de les mostres anteriors emmagatzemades en una memòria. Òbviament aquest promedi està en finestrat, tot i que no es perceben els seus efectes.

En resum, gràcies a la propietat d'ergodicitat és possible realitzar una optimització en mitjana sense haver de conèixer prèviament les característiques del canal. Això obre la porta a que l'algoritme es pugui implementar en multituds de canals possibles.



B Apèndix B. Publicacions

La recerca duta en aquest projecte ha permès escriure vàries publicacions per congressos i revista. A continuació, es mostren les publicacions pels congressos. La primera es va presentar al congrés *International ITG Workshop on Smart Antennas 2009* celebrat a Berlín. La segona es té prevista presentar-la al simposi *IEEE Globecom 2009 Wireless Communications Symposium*, en cas d'aprovar-se.

A LOW COMPLEXITY SPACE-FREQUENCY MULTIUSER SCHEDULING ALGORITHM

Ana I. Pérez-Neira^{1,2}, Pol Henarejos¹, Velio Tralli³, Miguel A. Lagunas^{1,2}

email: anuska@gps.tsc.upc.edu, pol@redyc.com, vtralli@ing.unife.it, m.a.lagunas@cttc.es
 (1) Dept. of Signal Theory and Communications - Universitat Politècnica de Catalunya (UPC) – Spain
 (2) Centre Tecnològic de Telecomunicacions de Catalunya (CTTC) - Spain
 (3) ENDIF - Engineering Department University of Ferrara - CNIT - Italy

ABSTRACT

This work presents a resource allocation algorithm in K-user, M-subcarrier and NT-antenna systems for on-line scheduling. To exploit temporal diversity and to reduce complexity, the ergodic sum rate is maximized instead of the instantaneous one. Dual optimization is applied to further diminish complexity together with a stochastic approximation, which is more suitable for online algorithms. Weighted sum rate is considered so that users can be either prioritized by higher layers or differentiated by proportional rate constraints. The performance and complexity of this algorithm is compared with well-known benchmarks and also evaluated under real system conditions for the MIMO Broadcast channel.

1. INTRODUCTION

In a Multi-User MIMO (MU-MIMO) spatial multiplexing scheme, multiple users are scheduled in the same resource block. MU-MIMO is a promising way to increase system throughput and there is a growing interest on the topic as [1,2, 3, 4] shows. Recently attention has been paid to the combination of spatial diversity multiple access systems and frequency domain packet scheduling [5,6,7,8,9]. Specifically, in [6] the authors present a low complexity sum-power constraint iterative waterfilling that is capacity achieving. It improves the convergence of [3] and is probably convergent. In [8] the authors address the problem of feedback reduction. The present paper aims at both, low complexity and reduced feedback. In contrast to [7], in order to further reduce complexity for on line implementation we follow a dual decomposition strategy and a stochastic approximation. In order to reduce feedback load the paper resorts to opportunistic strategies that solve the spatial scheduling. More specifically, an efficient algorithm for optimal beam subset and user selection is performed to find the best trade-off between the multiplexing gain and the multiuser interference in the opportunistic scheme when the number of users is not high. In summary, this paper proposes a joint spatial and frequency scheduler that allows on-line implementation and only requires partial or low feedback and a low-complexity implementation.

This paper is organized as follows. The space-frequency scheduler is formulated in Section 2 and the distributed scheme that is proposed based on dual optimization is presented in Section 3. Section 4 explains the low complexity ergodic algorithm, together with an evaluation of its

complexity. The numerical results are presented in Section 5, and Section 6 concludes the paper.

2. PROBLEM FORMULATION

We consider an OFDMA scenario with M subcarriers and K users. Each user k is single antenna and receives simultaneously N_T signals, which can come from different spatial locations, antennas or beams. Only one of the N_T signals is intended for user k . The received signal by user k on the subcarrier m is given by

$$y_{k,m} = a_{k,m} \sqrt{p_{k,m}} \alpha_{k,m,k_q} s_{k,m} + \sum_{s \neq k} a_{s,m} \sqrt{p_{s,m}} \alpha_{k,m,s_q} s_{s,m} + w_{k,m} \quad (1)$$

where, $\mathbf{a} = \{a_{k,m}\}$ is the set of binary allocation variables, i.e.

$a_{k,m} = 1$ if user k is scheduled on frequency m , $a_{k,m} = 0$ otherwise, and $\mathbf{p} = \{p_{k,m}\}$ is the set of allocated powers, $s_{k,m}$ is the information signal of user k through frequency m , $E\{|s_{k,m}|^2\} = 1$. Finally, α_{k,m,k_q} denotes the equivalent channel seen by the k th user at frequency m with respect to the q th beam, antenna or transmitter associated with user k . For instance, in the case of a MISO (Multiple input single output) broadcast channel, $|\alpha_{k,m,k_q}|^2 = c_{k,m,k_q} = |\mathbf{h}_{k,m}^T \mathbf{b}_{m,k_q}|^2$, where

\mathbf{b}_{m,k_q} is the beamforming vector that is associated with user k and that is obtained from the set of beams $k_q \in \{1, \dots, N_T\}$. In this case the number of interference terms in (1) is equal to $N_T - 1$. From a viewpoint of information theory, the model in (1) could correspond either to a broadcast channel or to an interference channel with N_T transmitters and K receivers. In spite of the big gains in spectral efficiency that can be obtained by incorporating multiantenna transmission to a multicarrier system, an evident drawback of this scenario is the increased design complexity. In other words, multiantenna, multiuser and multicarrier channels significantly increase the set of design parameters and degrees of freedom at the PHY layer. In this work, the focus is on the optimization of the PHY layer parameters with low complexity burden. Concerning the optimality criteria, we consider the problem of rate maximization in (2) with power constraints and also proportional rate constraints.

$$\begin{aligned}
\max_{\mathbf{a}, \mathbf{p}} R(\mathbf{a}, \mathbf{p}) &= E_{\gamma} \left\{ \sum_{k=1}^K \sum_{m=1}^M \log_2 \left(1 + \gamma_{k,m,k_q}(\mathbf{a}, \mathbf{p}) \right) \right\} \\
s.t. \quad E_{\gamma} \left\{ \sum_{m=1}^M \log_2 \left(1 + \gamma_{k,m,k_q}(\mathbf{a}, \mathbf{p}) \right) \right\} &\geq \phi_k R, \quad k=1, \dots, K \\
\sum_{k=1}^K \phi_k &= 1 \\
E_{\gamma} \left\{ \sum_{m=1}^M \sum_{k=1}^K a_{k,m} p_{k,m} \right\} &\leq \bar{P} \\
a_{k,m} &\in \{0, 1\} \quad m=1, \dots, M \\
p_{k,m} &\geq 0 \quad m=1, \dots, M \quad k=1, \dots, K
\end{aligned} \tag{2}$$

with

$$\gamma_{k,m,k_q}(\mathbf{a}, \mathbf{p}) = \frac{a_{k,m} p_{k,m} c_{k,m,k_q}}{\sigma^2 + \sum_{s \neq q} a_{s,m} p_{s,m} c_{k,m,k_s}} \tag{3}$$

where \mathbf{a} and \mathbf{p} are vectors whose components are $a_{k,m}$ and $p_{k,m}$, respectively. $\gamma_{k,m,k_q}(\mathbf{a}, \mathbf{p})$ is the SINR (Signal to Interference and Noise Ratio) of user k at frequency m and associated with beam q and c_{k,m,k_q} denotes the equivalent channel power gain seen by the k th user at frequency m with respect to the q th beam, antenna or transmitter. The formulation of the sum rate in (2) indicates that at each frequency up to N_T transmissions can be spatially multiplexed. We assume that $\gamma_{k,m,k_q}(\mathbf{a}, \mathbf{p})$ are known by the N_T transmitters by means of partial channel feedback. For instance, this would be the case of a broadcast channel where the Base Station (BS) has perfect SINR feedback. Other possible scenario is that of N_T BS's in a cellular system; in this case the assumption would be that all BS know the equivalent channel magnitude. ϕ_k are the weights that allow prioritizing the users. The problem to solve deals with scheduling of users and powers, the spatial precoder is fixed and part of the initial conditions of the problem.

Rate optimization is a reasonable choice for utility, reflecting the various coding rates implemented in the system. We assume an idealized link adaptation protocol. Proportional rate constraints allow a more definitive prioritization among the users, which is quite useful for service class differentiation. Theoretically, this formulation also traces out the boundary of the capacity region similar to the weighted sum-rate maximization. The main difference is that it actually identifies the points on the capacity region boundary that satisfy the rate proportionally constraints. Furthermore, the max-min rate formulation is a special case of this formulation, i.e., when $\phi_1 = \dots = \phi_K$. Finally, by enforcing the average power constraint we allow instantaneous power levels to exceed the average power when necessary. Sum power constraint is needed in scenarios such as BC channel, but it is not usually imposed in multi-cell scenarios.

Note that ergodic optimization is considered because of twofold: i) it reduces the complexity of the resulting algorithm and ii) it incorporates the time dimension in the resulting resource allocation. In other words, in the case of instantaneous rate allocation only, the OFDMA algorithms

are re-run every symbol (or several symbols). In this paper, we can capture the idea of "time slot allocation" by using the ergodicity assumption, and determine power allocation functions that are parameterized by the channel knowledge.

Note also that if there is no frequency structured components or noise, the maximal sum rate signaling does not require introducing correlation between subcarriers (cooperative subcarrier transmission or joint frequency-space processing). Therefore, the problem is separable across the subcarriers, and is tied together only by the power constraint. In these problems, it is useful to approach the problem using duality principles. In addition, the utility function is non convex and by solving the dual problem and formulating it as a canonical distributed algorithm [10], the algorithm is simplified and also convergence to the globally optimal rate allocation can be achieved.

Finally and as notational convention vectors are set in boldface.

3. DUAL OPTIMIZATION

The proposed algorithm is based on a dual optimization framework. In other words, it is based on a Lagrangian relaxation of the power constraints and (possibly) rate constraints. This relaxation retains the subcarrier assignment exclusivity constraints, but "dualizes" the power/rate constraints and incorporates them into the objective function, thereby allowing us to solve the dual problem instead. This dual optimization is much less complex as we explain next. To derive the dual problem we first write the Lagrangian. In order to simplify notation we define

$$\begin{aligned}
r_k &\triangleq \sum_{m=1}^M E_{\gamma} \left\{ \log_2 \left(1 + \gamma_{k,m,k_q} \right) \right\} \\
\hat{p}_k &\triangleq \sum_{m=1}^M E_{\gamma} \left\{ a_{k,m} p_{k,m} \right\}
\end{aligned} \tag{4}$$

where we do not explicitly write the dependence of r_k and \hat{p}_k on the optimization variables \mathbf{a} , \mathbf{p} . Based on this definitions, the Lagrangian is

$$L = R(1 - \boldsymbol{\mu}^T \boldsymbol{\phi}) + \lambda \bar{P} - \lambda E_{\gamma} \left\{ \sum_{k=1}^K \hat{p}_k \right\} + \sum_{k=1}^K \mu_k r_k \tag{5}$$

λ , $\boldsymbol{\mu}$ are the dual variables (also called prices) that relax the cost function, $R = \sum_{k=1}^K r_k$ and \bar{P} is the power constraint.

Focusing on the first term in the maximization, we observe that if $(1 - \boldsymbol{\mu}^T \boldsymbol{\phi}) > 0$ then the optimal solution would be

$R^* = \infty$, since R is a free variable. This is clearly an infeasible solution for ergodic sum rate. Furthermore, if $(1 - \boldsymbol{\mu}^T \boldsymbol{\phi}) < 0$ then the optimal solution would be $R^{opt} = 0$.

Thus, we would like to constrain the multiplier to satisfy $\boldsymbol{\mu}^T \boldsymbol{\phi} = 1$. Thus, (5) can be simplified to

$$L = \lambda \bar{P} - \lambda E_{\gamma} \left\{ \sum_{k=1}^K \hat{p}_k \right\} + \sum_{k=1}^K \mu_k r_k \tag{6}$$

Note that the weights μ_k are the dual multipliers that enforce the proposed rate constraints. Additivity of the utility and

linearity of the constraints lead to the following Lagrangian dual decomposition into individual user terms

$$L = \sum_k L_k + \lambda \bar{P} \quad (7)$$

where, for each user k ,

$$L_k = \sum_k \mu_k r_k - \lambda_k \sum_k \hat{p}_k \quad \lambda_k = \lambda \quad (8)$$

only depends on local rate r_k and the prices λ, μ . The dual function $g(\lambda, \mu)$ is defined as:

$$\begin{aligned} g(\lambda, \mu) &= \max_{\mathbf{a}, \mathbf{p}} \left(\sum_k L_k(\mathbf{a}, \mathbf{p}, \lambda_k, \mu) + \lambda \bar{P} \right) = \\ &= \sum_k L_k^*(\mathbf{a}^*, \mathbf{p}^*, \lambda_k, \mu) + \lambda \bar{P} \end{aligned} \quad (9)$$

Evidently, this dual problem involves only $K+1$ variables and it is, therefore much easier to solve than the primal problem. Moreover, the maximization in (9) can be conducted parallel by each user, as long as the aggregate link price λ is feedback to source user k . Note that if there were no global constraint, as it is the case in multicell systems where $p_{k,m}$ stands for the power that Base Station k has to allocate in frequency m and there is only per BS power constraints, the problem is further simplified.

The dual problem is defined as:

$$\begin{aligned} \min g(\lambda, \mu) \\ \text{s.t. } \lambda \geq 0, \mu \in D \quad D = \{\mu \geq 0, \mu^T \phi = 1\}. \end{aligned} \quad (10)$$

Since $g(\lambda, \mu)$ is the pointwise supremum of a family of affine functions in λ, μ , it is convex and (10) is a convex minimization problem (even if the primal is not a concave maximization problem).

Since $g(\lambda)$ may be non differentiable, an iterative subgradient method can be used to update the dual variable λ to solve the dual problem. The computation of the subgradient requires knowing the individual weighted ergodic rates per user. Note that the “weights” in this case are no longer predetermined constants, but are effectively the multipliers that enforce the proportional rate constraints. From an initial guess λ^0 and μ^0 , the subgradient method generates a sequence of dual feasible parts according to the iteration

$$\lambda^{i+1} = [\lambda^i - s^i g_\lambda^i]^+ \quad \mu^{i+1} = \Pi_D[\mu^i - s^i \mathbf{g}_\mu^i] \quad (11)$$

Where g_λ^i denotes the subgradient of $g(\lambda^*(\mu^i), \mu^i)$ with respect to λ

$$g_\lambda^i = \bar{P} - E_\gamma \left\{ \sum_k \hat{p}_k^* \right\} \quad (12)$$

and s^i is a positive scalar step-size. \mathbf{g}_μ^i denotes the subgradient of $g(\lambda^*(\mu^i), \mu^i)$ with respect to μ

$$\begin{aligned} g_\mu^i &= \bar{\mathbf{R}}^i - \phi \bar{\mathbf{R}}^i \\ \text{with } [\bar{\mathbf{R}}^i]_k &= \sum_m E_\gamma \left\{ \log_2 \left(1 + \gamma_{k,m,k_q}(\mathbf{a}^*, \mathbf{p}^*, \lambda^i, \mu_k^i) \right) \right\} \\ \bar{\mathbf{R}}^i &= \sum_k [\bar{\mathbf{R}}^i]_k \end{aligned} \quad (13)$$

Finally, $\Pi_D[\cdot]$ denotes projection onto the set D .

Concerning convergence, for a primal problem that is a convex optimization, the convergence is towards a global optimum. Otherwise, global maximum of non concave functions is an intrinsically difficult problem on non convex optimization. In [10] the authors show that the sequence of the maximization of (8) and the computation of (11) forms a canonical distributed algorithm that solves (2) and the dual problem (10). Even for non concave utilities the canonical distributed algorithm may still converge to a globally optimal solution if L_k^* is continuous at optimum λ^* . Based on this property, an analytical proof of convergence for the algorithm that is proposed next is left for further work. Simulation results have proved good convergence for it.

4. ALGORITHM

In the rest of the paper we deal with the specific case of Broadcast (BC) channel, where the global power constraint is needed. The maximization of (9) could have been formulated only with respect to the powers, \mathbf{p} . In this way, whenever any of the optimal components $p_{k,m}^* = 0$, this would mean that user k should not be scheduled in frequency k . In addition to the complexity of this multiuser frequency power allocation problem, note that $p_{k,m}$ depends on the spatial channel at frequency m , which, in the case of the BC channel, depends on the spatial precoder that is associated with user k . The purpose of this work is to design a low complexity scheduler; this fact motivates the simplification of the complex space-frequency multiuser scheduler that has been described by using the multibeam opportunistic scheme [11]. In this case, the Base Station uses a set of orthonormal beams that are associated with the users depending on their reported SINR. This scheduler is designed such that it works without interacting with power allocation and dual optimization. The problem formulation of (9) accounts for this explicit user scheduling by incorporating the discrete variables \mathbf{a} . As shown in Fig. 1, the first step in the proposed algorithm is the spatial scheduler, which obtains \mathbf{a} in a low complexity way, as it is described in Section A.



Fig. 1. Block diagram of the proposed algorithm.

Once user k has been selected by being associated with beam q at each frequency m , the next step in the algorithm (see Fig. 1) is the power allocation, which should be derived from the equation

$$\frac{\partial L_k}{\partial p_{k,m}} = 0 \quad \forall k \quad (14)$$

Assuming the interference with constant power \bar{P} , a suboptimal solution is provided by the waterfilling in (15)

$$p_{k,m}^* = \left[\frac{\mu_k}{\lambda \ln(2)} - \frac{1}{\gamma_{k,m,k_q}^1} \right]^+ \quad (15)$$

where

$$\gamma_{k,m,k_q}^1 = \frac{c_{k,m,k_q}}{\sigma^2 + \sum_{s \neq q}^{N_T} a_{s,m} P c_{k,m,k_s}} \quad (16)$$

The dual variables are obtained by solving the dual problem (10) with low complexity in a stochastic way. The detailed waterfilling procedure is described in Section C.

A. Spatial scheduler

The spatial scheduler is obtained for each frequency m by introducing a simplified SINR that considers uniform power allocation (i.e. no interaction neither with the primal variables \mathbf{p} nor with the dual variables λ, μ). The proposed spatial scheduler is based on the multibeam opportunistic strategy, which considers an orthonormal random beamforming set as precoder and assigns users to beams based only on SINR feedback. In order to counteract the losses that opportunistic schemes present when the number of users is moderate or low we extend the beam and user set optimization proposed in [12] to OFDMA. Let $Q_m \in \{1, \dots, N_T\}$ denotes the number of users served or active beams at m th frequency bin, $U_m^{(Q_m)} \subset \{1, \dots, K\}$ and $S_m^{(Q_m)} \subseteq \{1, \dots, N_T\}$ are the user and beam set, respectively, with Q_m elements without repetition. For the BC the SINR in (3) is equivalent to

$$\gamma_{k,m,q}^{(Q_m)} = \frac{p_{k,m} |\mathbf{h}_{k,m}^T \mathbf{b}_{m,q}|^2}{\sigma_n^2 + \sum_{\substack{q \neq q \\ q \in S_m^{(Q_m)}}} p_{q,m} |\mathbf{h}_{k,m}^T \mathbf{b}_{m,q}|^2} \quad m \in S_m^{(Q_m)} \quad k \in U_m^{(Q_m)} \quad (17)$$

where $\mathbf{b}_{m,q}$ is q th beam at m th frequency and $\sigma_n^2 = \sigma_w^2/P$ is the noise variance. In (17) there are $\binom{K}{Q_m} Q_m!$ permutations of

$\gamma_{k,m}^{(Q_m)}$ for $S_m^{(Q_m)}$. The optimal beam subset S_m^* and user subset U_m^* at m th frequency bin are obtained by (9) using exhaustive search. However we apply a suboptimal approach to reduce the complexity. Next, we assume that equal power P is allocated among beams and subcarriers, $j=1, \dots, \binom{N_T}{Q_m}$ is the

index of all beam combinations.

Fixing j th combination, the optimal user satisfies

$$k_{m,j,q}^* = \arg \max_k \gamma_{k,m,j,q}^{(Q_m)} \quad (18)$$

where subindex k_q in (3) has become indexes k,j,q in (18) due to the beam and user search. The optimal value of (18) is added to $U_{j,m}^{(Q_m)}$. For each frequency m , optimal j th index is found by expression

$$j^* = \arg \max_{1 \leq j \leq \binom{N_T}{Q_m}} \sum_{q=1}^{Q_m} \log_2 \left(1 + \gamma_{k_{m,j,q}^*, m, j, q}^{(Q_m)} \right) \quad (19)$$

Hence, the optimal user and beam set that serves Q_m users simultaneously are $U_m^{*(Q_m)} = U_{j^*,m}^{(Q_m)}$ and $S_m^{*(Q_m)} = S_{j^*,m}^{(Q_m)}$. Finally,

optimal Q_m is given by

$$Q_m^* = \arg \max_{1 \leq Q_m \leq N_T} \sum_{k \in U_m^{(Q_m)}} \sum_{q \in S_m^{(Q_m)}} \log_2 \left(1 + \gamma_{k,m,j^*,q}^{(Q_m)} \right) \quad (20)$$

Once this spatial scheduling is finished for each frequency m , we simplify indexes $\binom{Q_m^*}{k_{m,j^*,q}^*, m, j^*, q} \triangleq k, m, k_q$ as they are used in

(3), ready to be applied in the frequency waterfilling and dual optimization step. The corresponding user selection variable $a_{k,m}$ is set to 1 if the user has been chosen.

Regarding fairness, this solution has the drawback of scheduling users on the available beams and frequencies by only looking at channel gains. In this way, users with good channel conditions, i.e. users located near the base station with a small path loss, tend to monopolize channel resources. The lack of resources for weak users may constrain the behavior of dual optimization. An improved spatial scheduler can be designed by considering the maximization of

$$\max_{k_q} \sum_k \sum_q \mu_k \log_2 (1 + \gamma_{k,m,k_q}) \quad \forall m \quad (21)$$

by using the equal power approximation of $p_{k,m}$, i.e. $p_{k,m} = P$. This can be simply implemented by inserting μ_k in (19) and (20). This scheduler interacts with the dual optimization algorithm. It releases beams and frequencies to users according to rate constraints, but preserves both the light requirements on feedback parameters and the distributed implementation.

Finally, instead of MOB other spatial precoders (such as Zero Forcing) can be used, however, at the expense of complexity increase in the feedback.

B. Feedback complexity

The number of feedback parameters to be transmitted to the BS by each k th user are $N_T \times M$ and they correspond to all possible values of $|\mathbf{h}_{k,m}^T \mathbf{b}_{m,q}|^2$ for $q=1, \dots, N_T$ and $m=1, \dots, M$.

This amount can be reduced by fixing Q_m previously. Users must know the value Q_m . In this case, feedback is reduced to $3 \times M$ and they correspond to q^*, j^* , best beam and best permutation, and γ_{k,m,j^*,q^*} . That is because fixing Q_m causes the number of permutations is known a priori by all users and it is fixed. Thus, each user can compute γ_{k,m,j^*,q^*} and send it to BS, jointly with indexes. In addition, fixing $Q_m = Q = N_T$ the amount of feedback is also reduced to $2 \times M$, q^* and $\gamma_{k,m,1,q^*}$ since $j=1$.

Finally, depending on the delay spread of the channel, the number of parameters to feedback can be further reduced by frequency grouping or chunk processing [5].

C. Frequency power allocation

Once the user selection problem has been solved, \mathbf{a} is known in (2), and the next step is to obtain the solution for the power allocation \mathbf{p} in (12).

To compute (12), ergodic maximization through stochastic approximation is introduced [14]. In other words, maximization occurs through time or iterations. At n th iteration, power assigned to k th user at m th carrier becomes

$$p_{k,m}^*[n] = \left[\frac{\mu_k[n]}{\lambda[n] \ln(2)} - \frac{1}{\gamma_{k,m,k_q}^1[n]} \right]^+ \quad (22)$$

where $[x]^+ = \max(x, 0)$.

Per-user rate and total power are given by

$$\begin{aligned} R_k[n] &= \sum_{m \in M} R_{k,m} \left(\gamma_{k,m,k_q} (p_{k,m}^*[n]) \right) \\ P[n] &= \sum_{m \in M} \sum_{k \in K} p_{k,m}^*[n] \end{aligned} \quad (23)$$

Finally, λ, μ are updated using subgradient method as in (11)-(13) but with a stochastic approximation. In fact, these parameters are given by expressions

$$\begin{aligned} \lambda[n+1] &= [\lambda[n] - \delta(\bar{P} - P[n])]^+ \\ \mu[n+1] &= \Pi_{\mathcal{D}}(\mu[n] - \delta(R[n] - \phi R[n])) \end{aligned} \quad (24)$$

where $\Pi_{\mathcal{D}}$ is the projection onto set $\mathcal{D} = \{\mu \geq 0 | \mu^T \phi = 1\}$,

$\mathbf{R}[n] = [R_1[n] \ \cdots \ R_K[n]]^T$ and $R[n] = \sum_{k \in K} R_k[n]$. To

obtain a good performance, a suitable values could be $\lambda[0]=1$, $\mu[0] = \frac{\phi}{\phi^T \phi}$ and $\delta=0.01$.

Subgradient search methods have been used to obtain the solution for the dual variables. As it is important to be able to perform resource allocation in real-time, we obtain an on-line adaptive algorithm by performing the iterations of the subgradient across time.

D. Complexity

Using ergodic sumrate relaxes complexity since constraints are not instantaneous but ergodic. Note that though objective is ergodic, feedback parameters contain instantaneous information. This algorithm has several stages. First of all there is the pooling stage. During this step, BS sends a pilot signal through all beams and only one beam is active. The complexity in the beamforming is $O(MN_T)$.

Next step is computing all γ parameters. Its complexity depends on fixing Q_m . Leaving it as a free parameter, complexity is $O(MK2^{NT}N_T^2)$. Otherwise, fixing it, complexity becomes $O(MK2^{NT}N_T)$. Moreover, adjusting $Q_m = Q = N_T$ it is reduced to $O(MKN_T)$.

Finally, there is the power allocation stage, water-filling has complexity $O(MK)$ and it is followed by $O(K)$ updates for the rates, power and multipliers.

In general, complexity could be very low, as $O(MKN_T)$, or higher, as $O(MK2^{NT}N_T^2)$, depending on how optimum is desirable.

Next table shows algorithm step-by-step and its complexity.

Note that in step 2 there are three possibilities: 2.a has $N_T \times M$ parameters of feedback; 2.b, $3 \times M$ and 2.c, $2 \times M$.

Other solutions such as [13] find the optimal bound of capacity rate at cost of complexity, $O(K^2 M \log N)$. Others have less complexity, as $O(M)$ in [8], but they lose in performance. See fig. 4 for more details in the comparison.

5. RESULTS

We organize numerical results in two parts: the first part illustrated in section 1 refers to a simple cellular scenario and has the aim of showing the main behavior of the algorithm; the second part refers to a more realistic scenario.

1. Pooling: BS transmits pilot signal to sense each equivalent channel.	$O(MN_T)$
2.a <u>Non-fixed</u> Q_m Feedback: each user sends $ \mathbf{h}_{k,m}^T \mathbf{u}_{m,q} ^2$ BS schedules users spatially	$O(MN_T)$ $O(MK2^{NT}N_T^2)$
2.b <u>Fixed</u> Q_m Feedback: each user sends $q^*, j^*, \gamma_{k,m,j^*,q^*}$ BS schedules users spatially	$O(M)$ $O(MK2^{NT}N_T)$
2.c $Q_m = N_T$ Feedback: each user sends $q^*, \gamma_{k,m,1,q^*}$ BS schedules users spatially	$O(M)$ $O(MKN_T)$
3. Water-filling	$O(MK)$
4. Updating parameters λ, μ	$O(K)$

Table 1. Algorithm step-by-step and its complexity.

A. Results for simple scenario

All simulations in this scenario consider $M=64$ subcarriers, power constraint $\bar{P}=10\text{dB}$ and power parameter $P=1$ in equation (16). The channel model includes normalized Rayleigh fading and does not take care of path-loss or shadowing components. All carriers have frequency spacing of 1Hz. All users are located at same distance from BS. A linear array of N_T antenna is considered at the base station and distance between sensors is 0.5λ .

Fig. 2 shows how every user converges to its weight with few iterations. Note that user rates are normalized with sumrate. Fig. 3 shows power convergence to average power constraint. Hence the good convergence properties of dual optimization algorithm are confirmed by the results.

In order to compare the spatial scheduling algorithm with other algorithms of the literature, we show in Fig. 4 the rate region for two users, compared with those obtained with the algorithms in [13] (DPC) and [8] (KOU), and to uniform power allocation strategy (UPA). Note that [13] has higher complexity and fixes a theoretical maximum sum-rate that can be achieved, whereas [8] has lower complexity, but worse performance.

Fig. 5 shows the computation complexity in terms of simulation time for different system configurations and strategies of choosing Q_m . Note how complexity increases with the usage of dynamic Q_m . In order to compare the complexity with that of [13], based on DPC, we provide the following table that shows the computation complexity for $K=8$ and $K=32$ users, and $N_T=2$. Time is expressed in seconds. This gives a clear idea of the complexity of [13], $O(K^2 M \log N)$, and how it increases with number of users.

	K=8	K=32
[13]	283	3349
Our algorithm (Dyn Q_m / $Q_m=N_T$)	28 / 31	50 / 62

Table 2. Complexity comparison with [13].

When compared to solution [8], for $N_T=2$ and $K=2$, our algorithm requires roughly the same computation time, but achieves best performance, as shown in Fig. 4.

Finally, Fig. 6 shows the impact on sum-rate of using a dynamic Q_m or a fixed Q_m . Dynamic choice of Q_m is useful for a large number of antennas, whereas the other choice is good for few users.

B. Results in a realistic scenario

In this section we present and discuss simulation results obtained for a scenario which incorporate some characteristic aspects of practical application in next generation wireless systems. In fact, 3GPP-LTE and WiMAX employ OFDMA as their main multiple access mechanism (although other options are also defined in the standards).

We are considering here a single cell of the downlink of an OFDM wireless system with $M=128$ subcarriers working on a bandwidth of 1.25 Mhz. Base station is equipped with multiple antennas. The system is TDD and it is assumed that 2/5 of frame interval is used for downlink transmission. The CSI coming from users is updated every 10ms. Two options for user distribution are considered: in the first option the users have a position which is uniformly distributed in circular area of radius 500m; in the second option the users are placed at the same distance of 250 m from the base station.

Channel model includes path loss, correlated shadowing (not present in the second option for user distributions) and time and frequency correlated fast fading. Path loss is modeled as a function of distance as $L(\text{db}) = k_0 + k_1 \log(d)$ ($k_1=40$, $k_1=15.2$ for results). Shadowing is superimposed to path-loss, with classical lognormal model ($\sigma=6$ dB) and exponential correlation in space (correlation distance equal to 20m). Fast fading on each link of the MIMO broadcast channel is complex Gaussian, independent across antennas and is modeled according to a 3GPP Pedestrian model [11]. This model has a finite number of complex multipath components with fixed delay (delay spread around 2-3 microseconds) and power (average normalized to 1). Time correlation is obtained according to a Jakes' model [12] with given Doppler bandwidth (6 Hz in the results). At the base station orthogonal beamforming is adopted, where beam vectors change randomly at each frame. In the simulated system the total average power constraint is fixed to 1W.

Note that, in realistic conditions, channel variations in time due to Doppler effects have a non negligible impact on the feedback quality. In fact, at the scheduling time n the algorithm uses feedback parameters measured at time $n-1$, which can be changed in the meanwhile. Therefore, due to outdated feedback the transmission at the scheduled rate may fail sometimes. This aspect is left for future investigation

In the first two figures, fig. 7 and fig. 8, the dynamic behavior of sum-rate and total allocated power is illustrated for a system with 10 users in fixed position at distance 250 m. In the scheduling algorithm equal weights $\phi_k = 1/K$ are used. We note that although total power and sum rate change frame by frame due to fast channel variations, the algorithm for dual variable optimization correctly tracks the constraint on the average power. We also observed from a wide set of results that the range of variations enlarges when the users have

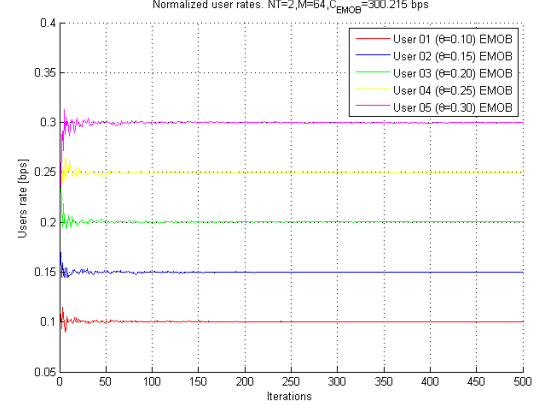


Fig. 2. Different rates for 5 users using this algorithm

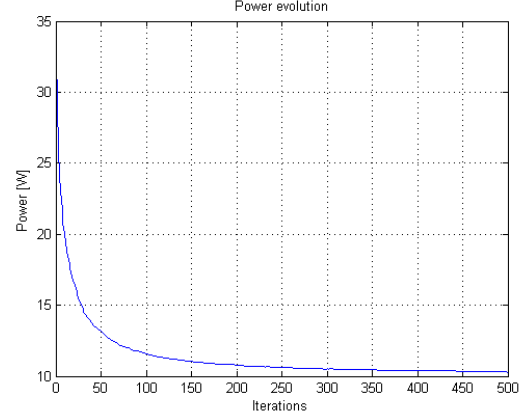


Fig. 3. Power evolution vs. iterations.

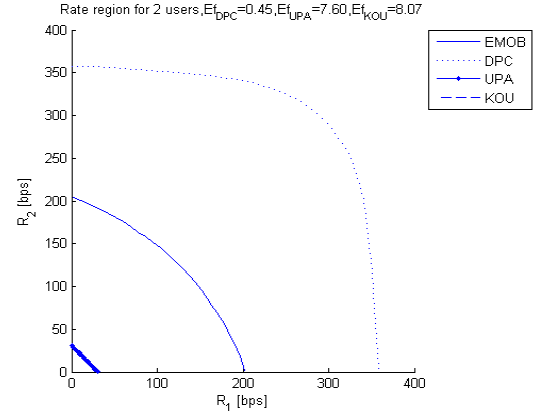


Fig. 4. Rate region for two users. DPC from [8], and KOU from [10] are plotted jointly with this algorithm EMOB. UPA is also plotted. Note that UPA and KOU are practically coincident.

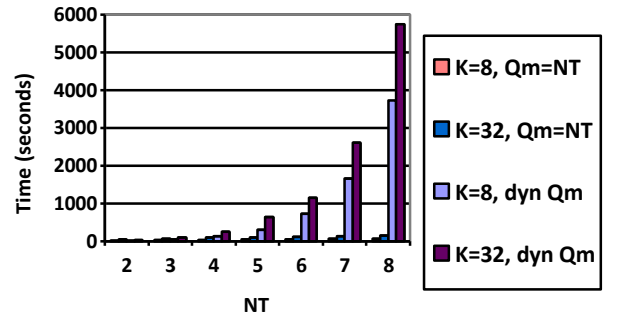


Fig. 5. Simulation time in seconds for $K=8$ and $K=32$, dynamic Q_m or fixed Q_m .

different path-loss, but again the algorithm tracks correctly the average.

In spite of rate and power variations, we also checked the robustness of algorithm to ensure fair average rate allocation among users. This is shown in the following two histograms in fig. 9 and fig. 10, which illustrates the distribution of average user rates in a system with 10 users, 3 antennas and 3 different user classes (class 1: weight $0.5/K$ - class 2: weight $1/K$ - class 3: weight $1.5/K$). Fig. 9 refers to equal distance users at 250 m, whereas fig. 10 refers to uniformly distributed users in a circular area of radius 500 m. We can note that the algorithms are quite fair to assign rates to different users, even when they have with different weights and different path-loss conditions.

Finally, fig. 11 shows sum rate and user rate vs. number of user, in a system with 3 antennas, users at equal distance 250 m from BS, and 3 different user classes (class 1: weight $0.5/K$ - class 2: weight $1/K$ - class 3: weight $1.5/K$). We note that sum rate increases with the number of users, meaning that the scheduling algorithm capture the available multiuser diversity while preserving average rate fairness. Per-user rate decreases since the sum-rate needs to be shared among an increasing number of users.

6. CONCLUSIONS

This paper has presented a low complexity space-frequency scheduler that allocates power among users. Ergodic objective and ergodic constraints are purposed to relax complexity. Moreover many strategies had been presented and low complexity has been explained. In addition, weights are purposed in order to set rate priorities or several QoS. Finally, some benchmarks are presented to compare the performance. Aspects such as robustness to imperfect CSIT, discrete rate allocation, modification of the algorithm to incorporate jointly encoded sub-channels (e.g. space-time codes) and cross-layer design for user scheduling improvement are possible topics for further research. The space-frequency multiuser scheduler has been presented in a general formulation such that the proposed distributed strategy (as a result of the dual optimization and opportunistic user selection) together with the low complexity of the proposed ergodic scheduler can be applied to different space-frequency scheduling scenarios.

7. REFERENCES

- [1] J.Brehmer, W. Utschick, "Nonconcave Utility Maximisation in the MIMO Broadcast Channel," *Eurasip JASP*, vol. 2009.
- [2] M. Haardt, V. Stankovic, G. Del Galdo, "Efficient Multi-user MIMO Downlink Precoding and Scheduling," *CAMSAP* 05, Dec. 2005.
- [3] Sriram Vishwanath, Nihar Jindal, and Andrea Goldsmith, "On the Capacity of Multiple Input Multiple Output Broadcast Channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1570–1580, Apr. 2005.
- [4] W. Yu, "Uplink–Downlink Duality Via Minimax Duality," *IEEE Trans. on IT*, February 2006.
- [5] E. Jorswieck, A. Sezgin, B. Ottersten, A. Paulraj, "Feedback reduction in uplink MIMO OFDM Systems by Chunk Optimization," *Eurasip JASP*, Vol. 2008.

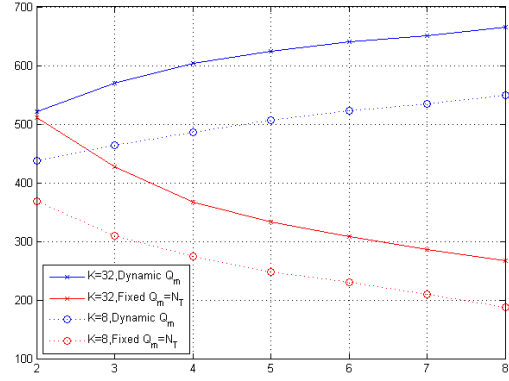


Fig. 6. Contribution to sum-rate of dynamic Q_m or fixed Q_m , plotted for $K=8$ and $K=32$ users with different number of antennas.

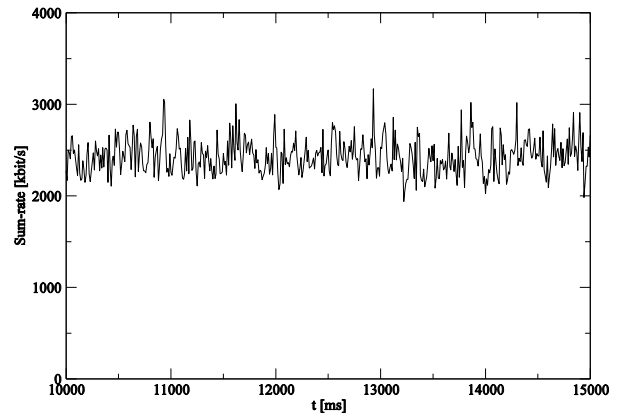


Fig. 7. Dynamic behavior of sum-rate in a cell with 10 users in fixed positions at distance 250m from BS.

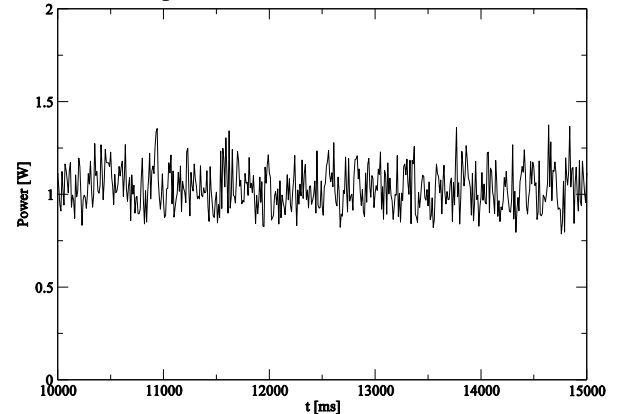


Fig. 8. Dynamic behavior of total allocated power in a cell with 10 users in fixed positions at distance 250m from BS.

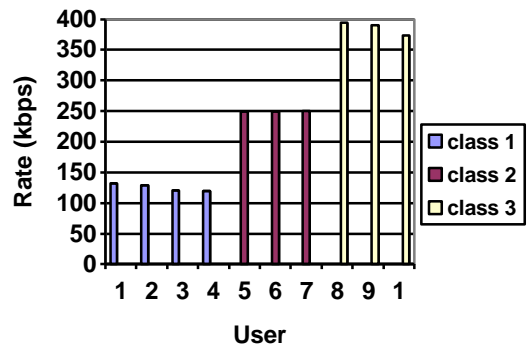


Fig. 9. Per-user average rate distribution in a system with 10 users at equal distance 250 m from BS, belonging to 3 different user classes.

- [6] N. Wei, A. Pokhariyal, T. B. Sorensen, T. E. Kolding, P.E. Mogensen, IEEE JSAC, vol.26, no. 6, Aug. 2008.
- [7] M. Codrenau, M. Juntti, M. Latva-Aho, "Low-complexity iterative algorithm for finding the MIMO-OFDM broadcast channel sum capacity", IEEE Trans. on Comm., vol. 55, no. 1, jan 2007.
- [8] Issam Toufik, Marios Kountouris, "Power allocation and feedback reduction for MIMO-OFDMA opportunistic Beamforming," VTC Spring 2006: 2568-2572
- [9] G. Liu; J. Zhang; F. Jiang; W. Wang, "Joint Spatial and Frequency Proportional Fairness Scheduling for MIMO OFDMA Downlink," Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on Volume , Issue , 21-25 Sept. 2007 Page(s):491 - 494
- [10] M. Chiang, S. Zhang, P. Hande, "Distributed rate allocation for inelastic flows: optimization frameworks, optimality conditions, and optimal algorithms," IEEE JSAC, vol. 23, no.1, Jan. 2005.
- [11] M.Sharif, B.Hassibi, "On the capacity of Broadcast Channels with Partial Side Information," IEEE Trans. on IT, Feb. 2005.
- [12] T. Kang and H. Kim, "Optimal Beam Subset and User Selection for Orthogonal Random Beamforming", IEEE Communications letters, vol. 12, no. 9, Sep. 2008.
- [13] M. Kobayashi, G. Caire, "An Iterative Water-Filling Algorithm for Maximum Weighted Sum-Rate of Gaussian MIMO-BC," IEEE JSAC, vol. 24, Aug. 2006.
- [14] I. Wong, B. Evans, "Optimal OFDMA resource allocation with linear complexity to maximize ergodic weighted sum capacity," ICASSP 07, Hawaii.

8. ACKNOWLEDGMENTS

This work was supported by the European Commission under project NEWCOM++ (216715), Optimix (Grant Agreement 214625) and by Spanish Government TEC2008-06327-C03-01. The work has been done during the 6 months stay of A. Perez-Neira at ACCESS/Signal Processing Lab, KTH (Stockholm).

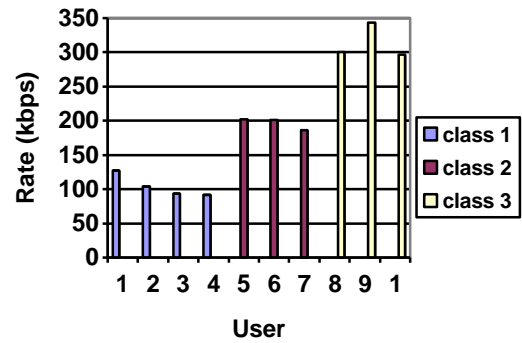


Fig. 10. Per-user average rate distribution in a system with 10 users uniformly distributed in a cell of radius 500 m, belonging to 3 different user classes.

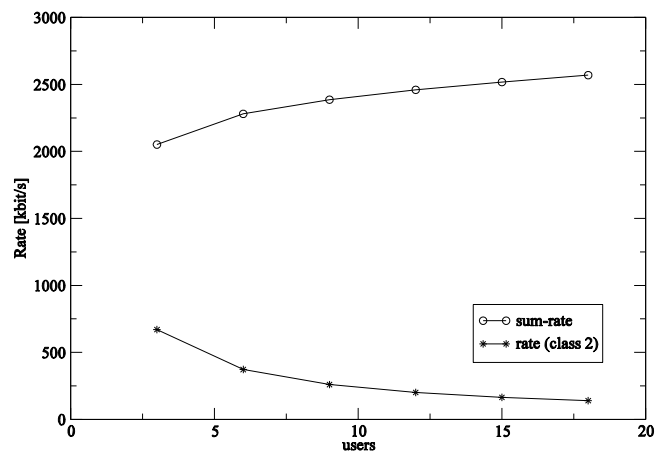


Fig. 11. Sum rate and per-user rate (user in class 2) vs. number of users, for a system with users at equal distance from BS.

A cross-layer multiuser scheduling strategy for heterogeneous traffic in the downlink of a MIMO-OFDMA systems

Velio Tralli *, Ana Perez-Neira #,§, Pol Henarejos #

Universitat Politècnica de Catalunya (UPC)- Barcelona - Spain

§ *Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)- Barcelona - Spain*

* *ENDIF, University of Ferrara - CNIT -Italy*

Abstract—In this paper we propose a cross-layer multiuser scheduling strategy for the support of heterogeneous traffic in the downlink of a MIMO-OFDMA system. It jointly considers different objectives: maximize the sum-rate on the radio channel, ensure a fair allocation of resources among users belonging to the same traffic class, consider the dynamics of traffic sources by limiting the delay of data packets in the queues, contribute to maximize quality of service figures at the application level. To exploit temporal diversity and to reduce complexity, the ergodic weighted sum-rate is maximized and dual optimization with stochastic approximation is applied to derive on-line algorithms. The numerical results show the capability of the scheduler to allocate physical layer resources according to rate constraints imposed for each different traffic class and with fairness inside each class, even in presence of different channels conditions and different network loads.

I. INTRODUCTION

In the development of fourth generation wireless networks, OFDM and MIMO technologies will be heavily exploited to support the transmission of multiple users. In fact, Multi-User MIMO (MU-MIMO) spatial multiplexing schemes are a promising way to increase system throughput and there is a growing interest on the topic as [1,2,3,4] shows. Recently, attention has been paid to the combination of spatial diversity multiple access systems and frequency domain packet scheduling [5,6,7,8,9]. Specifically, in [6] the authors present a low complexity sum-power constraint iterative waterfilling that is capacity achieving. It improves the convergence of [3] and is probably convergent. In [8] the authors address the problem of feedback reduction. On the other hand, future wireless networks need to provide connectivity to heterogeneous users offering different data traffic types, e.g. voice, video, web browsing, etc. This poses several constraints and additional challenges which can be faced within the framework of the cross-layer design [10,11].

In this paper we propose and investigate a cross-layer multiuser scheduling strategy for MIMO-OFDMA systems which jointly considers different objectives: maximize the sum-rate on the radio channel, ensure a fair allocation of resources among users belonging to the same traffic class, consider the dynamics of traffic sources by limiting the delay of data packets in the queues, contribute to maximize quality of service figures at the application level. In the proposed solution we look for a low computation complexity and a reduced feedback. In contrast to [7], in order to further reduce complexity for on line implementation we follow a dual decomposition strategy and a stochastic approximation. In order to reduce feedback load the paper resorts to opportunistic strategies that solve the spatial scheduling. In summary, this paper proposes a joint spatial and frequency scheduler that allows on-line implementation and only requires partial feedback and a low-complexity implementation.

The paper is organized as follows. After having introduces the radio interface model in Sect. II, the radio resource allocation problem is formulated in Sect. III. The dual optimization framework that allows on-line implementation is illustrated in Sect. IV and the allocation

algorithm follows in Sect. V. The architecture of the system applying the proposed solution is described in Sect. VI and, finally, results are presented in Sect. VII.

II. RADIO INTERFACE MODEL

We consider an OFDMA scenario with M subcarriers and K users. Each user k is single antenna and receives simultaneously up to N_T signals, which can come from different spatial locations, antennas or beams. Only one of the N_T signals is intended for user k . Each signal is characterized by a spatial signature, i.e. a beamforming vector $\mathbf{b}_{m,q}$, $m = 1, \dots, M$ $k = 1, \dots, N_T$. Subcarriers and spatial signatures are shared by all the users at each time slot. A binary allocation variable $\{a_{k,m,q}\}$ indicates whether a space-frequency resource is used by user k , i.e. $a_{k,m,q} = 1$ if user k is scheduled on frequency m and beam q , $a_{k,m,q} = 0$ otherwise. The composite signal received by user k on the subcarrier m is therefore given given by

$$y_{k,m} = \mathbf{h}_{k,m}^T \sum_{s \neq k} \sum_{q=1}^{N_T} a_{s,m,q} \sqrt{p_{m,q}} \mathbf{b}_{m,q} + \mathbf{w}_{k,m} \quad (1)$$

where $\mathbf{h}_{k,m}$ is the N_T -dimensional vector of channel gains, $p_{m,q}$ is the power of the q -th transmitted signal and $\mathbf{w}_{k,m}$ is the additive white Gaussian noise with variance σ_n^2 . From the viewpoint of information theory, the problem could correspond to a broadcast channel.

In spite of the big gains in spectral efficiency that can be obtained by incorporating multiantenna transmission to a multicarrier system, an evident drawback of this scenario is the increased design complexity. In order to keep feedback and computation complexity low in the optimization of the PHY layer parameters we consider in this paper opportunistic beamforming (OB) technique [12]. However, the scheduling strategy considered here can be also extended to a more general framework of spatial precoding. According to OB, the transmitter generate orthogonal spatial signatures randomly for each subcarrier. Based on partial CSI feedback, the scheduler and resource allocator only handle the set of binary allocation variables $\mathbf{a} = \{a_{k,m,q}\}$ and the set of powers $\mathbf{p} = \{p_{m,q}\}$. The CSI is the set of equivalent channel power gains $c_{k,m,q}$ seen by each user k at frequency m with respect to the q -th beam, which is given by $c_{k,m,q} = |\mathbf{h}_{k,m}^T \mathbf{b}_{m,q}|^2 / \sigma_n^2$.

III. RADIO RESOURCE ALLOCATION PROBLEM

The aim of resource allocation is to dinamically assign radio interface resources \mathbf{a} and \mathbf{p} to the different users in order to achieve the best tradeoff among different objectives:

- maximize the sum-rate on the radio channel
- ensure a fair allocation of resources among users belonging to the same traffic class and/or guarantee a minimum amount of resources to some users or classes

- consider the dynamics of traffic sources and limit the delay of data packets in the queue according to requirements of specific traffic classes.
- contribute to maximize quality of service figures at the application level

Motivated by the need of flexibility in considering the outlined objectives, we formulate the resource allocation problem at a given time instant or slot i^1 as a problem of weighted sum rate maximization with constraints on the average value of the rate provided to users as follows.

$$\max_{\mathbf{a}, \mathbf{p}} \sum_{k=1}^K \mathbb{E} \{w_k R_k(\mathbf{a}, \mathbf{p})\} \quad (2)$$

$$\begin{aligned} s.t. \quad & \mathbb{E} \left\{ \sum_{k=1}^K \sum_{m=1}^M \sum_{q=1}^{N_T} a_{k,m,q} p_{m,q} \right\} \leq P \\ & \mathbb{E} \{R_k(\mathbf{a}, \mathbf{p})\} = R_{0k}, \quad k \leq K_0 \\ & \mathbb{E} \{R_k(\mathbf{a}, \mathbf{p})\} \geq \phi_k \sum_{k=K_0+1}^K \mathbb{E} \{R_k(\mathbf{a}, \mathbf{p})\}, \quad k > K_0 \\ & \sum_{k=1}^K a_{k,m,q} \leq 1, \quad m = 1, \dots, M \quad q = 1, \dots, N_T \\ & p_{m,q} \geq 0, \quad m = 1, \dots, M \quad q = 1, \dots, N_T \end{aligned}$$

where

$$R_k(\mathbf{a}, \mathbf{p}) = \sum_{m=1}^M \sum_{q=1}^{N_T} \log_2(1 + \gamma_{k,m,q}(\mathbf{a}, \mathbf{p})) \quad (3)$$

is the rate provided to user k and

$$\gamma_{k,m,q}(\mathbf{a}, \mathbf{p}) = \frac{a_{k,m,q} p_{m,q} c_{k,m,q}}{1 + \sum_{s \neq k} \sum_{q=1}^{N_T} a_{s,m,q} p_{m,q} c_{s,m,q}} \quad (4)$$

is the SINR of user k at frequency m and beam q .

The coefficients w_k are the weights that allow prioritizing the users according to the service class and status of the queue buffers; they are not fixed constants, but are randomly changing parameters. The parameters R_{0k} , $k = 1, \dots, K_0$ and ϕ_k , $k = K_0 + 1, \dots, K$ are fixed parameters that define a constraint on the rate provided to users with

$$\sum_{k=K_0+1}^K \phi_k = 1 \quad (5)$$

Without loosing in generality, the first K_0 users have a fixed average rate constraint R_{0k} , whereas the remaining users have a proportional rate constraint depending on coefficients ϕ_k , which may be different for different traffic classes. We assume that $\gamma_{k,m,q}(\mathbf{a}, \mathbf{p})$ are known by the N_T transmitters by means of partial channel feedback. For instance, this would be the case of a broadcast channel where the Base Station has perfect SINR feedback.

It is important to underline that rate and sum power constraints are referred in this problem to average values. The reason is twofold: i) these constraints relax the instantaneous constraints leading to a reduction in the complexity of the resulting optimization algorithm and ii) it incorporates the time dimension in the resulting resource allocation by using the ergodicity assumption. For systems with hard instantaneous power constraint, the solution of the problem needs to be suitably adapted by using power rescaling, as shown later.

IV. DUAL OPTIMIZATION

The proposed algorithm is based on a dual optimization framework, based on a Lagrangian relaxation of power and rate constraints. This relaxation retains the subcarrier assignment exclusivity constraints,

¹the dependence on time i is omitted in the first formulation for the sake of clarity

but dualizes the power/rate constraints incorporating them into the objective function, thereby allowing us to solve the dual problem instead. This dual optimization is much less complex.

To derive the dual problem we first write the Lagrangian:

$$\begin{aligned} L(\mathbf{a}, \mathbf{p}, \lambda, \mu, \nu) = & \sum_{k=1}^K \mathbb{E} \{w_k R_k(\mathbf{a}, \mathbf{p})\} + \lambda P \\ & - \lambda \mathbb{E} \left\{ \sum_{k=1}^K \sum_{m=1}^M \sum_{q=1}^{N_T} a_{k,m,q} p_{m,q} \right\} + \sum_{k=1}^{K_0} \nu_k \mathbb{E} \{R_k(\mathbf{a}, \mathbf{p})\} - \sum_{k=1}^{K_0} \nu_k R_{0k} \\ & + \sum_{k=K_0+1}^K \mu_k \mathbb{E} \{R_k(\mathbf{a}, \mathbf{p})\} - \sum_{k=K_0+1}^K \mu_k \phi_k \sum_{k=K_0+1}^K \mathbb{E} \{R_k(\mathbf{a}, \mathbf{p})\} \end{aligned} \quad (6)$$

where the dual variables λ, μ, ν relax the cost function. It can be rewritten as

$$\begin{aligned} L(\mathbf{a}, \mathbf{p}, \lambda, \mu, \nu) = & \lambda P \\ & + \sum_{k=1}^{K_0} (\mathbb{E} \{w_k R_k(\mathbf{a}, \mathbf{p})\} + \nu_k (\mathbb{E} \{R_k(\mathbf{a}, \mathbf{p})\} - R_{0k}) - \lambda \mathbb{E} \{P_k(\mathbf{a}, \mathbf{p})\}) \\ & + \sum_{k > K_0} (\mathbb{E} \{w_k R_k(\mathbf{a}, \mathbf{p})\} + (\mu_k - \mu^T \phi) \mathbb{E} \{R_k(\mathbf{a}, \mathbf{p})\} - \lambda \mathbb{E} \{P_k(\mathbf{a}, \mathbf{p})\}) \end{aligned} \quad (7)$$

where $P_k(\mathbf{a}, \mathbf{p}) = \sum_{m=1}^M \sum_{q=1}^{N_T} a_{k,m,q} p_{m,q}$ is the power of user k .

The dual objective of problem (3) is defined as

$$g(\lambda, \mu, \nu) = \max_{\mathbf{a}, \mathbf{p}} L(\mathbf{a}, \mathbf{p}, \lambda, \mu, \nu) = L(\mathbf{a}^*, \mathbf{p}^*, \lambda, \mu, \nu) \quad (8)$$

where $\mathbf{a}^*, \mathbf{p}^*$ are the optimal solutions which maximize the Lagrangian for each feasible value of λ, μ, ν . Hence, the dual problem can then be written as

$$\begin{aligned} \min_{\lambda, \mu, \nu} \quad & g(\lambda, \mu, \nu) \\ s.t. \quad & \lambda \geq 0, \quad \mu \geq \mathbf{0} \end{aligned} \quad (9)$$

which is a convex optimization problem (even though the primal problem is not a concave maximization problem) with $K + 1$ variables. Since the dual objective may be not differentiable, an iterative subgradient method can be used to update the $K + 1$ solutions of the dual problem at time i . Starting from initial solutions λ^0 and μ^0, ν^0 , the update equations are:

$$\begin{aligned} \lambda^{i+1} &= [\lambda^i - \delta_\lambda g_\lambda^i]^+ \\ \mu^{i+1} &= [\mu^i - \delta_\mu \mathbf{g}_\mu^i]^+ \\ \nu^{i+1} &= [\nu^i - \delta_\nu \mathbf{g}_\nu^i] \end{aligned} \quad (10)$$

where g_λ^i is the subgradient of function $g()$ with respect to λ , i.e.

$$g_\lambda^i = P - \sum_{k=1}^K \mathbb{E} \{P_k(\mathbf{a}^*, \mathbf{p}^*)\}, \quad (11)$$

\mathbf{g}_μ^i and \mathbf{g}_ν^i are the subgradients of function $g()$ with respect to μ and ν , i.e.

$$\mathbf{g}_{\nu,k}^i = (\mathbb{E} \{R_k(\mathbf{a}^*, \mathbf{p}^*)\} - R_{0k}), \quad k \leq K_0$$

$$\mathbf{g}_{\mu,k}^i = \mathbb{E} \{R_k(\mathbf{a}^*, \mathbf{p}^*)\} - \phi_k \sum_{k=K_0+1}^K \mathbb{E} \{R_k(\mathbf{a}^*, \mathbf{p}^*)\}, \quad k > K_0 \quad (12)$$

and $\delta_\lambda, \delta_\mu, \delta_\nu$ are positive step-size parameters. An useful advantage of the iterative method is that it keeps low the computation complexity. In the practical implementation the evaluation of the average

power and rate at time i can be done through a stochastic approximation: if $f^i(\mathbf{a}^*, \mathbf{p}^*)$ is the generic rate or power at time i , its average $\mathbb{E}\{f^i(\mathbf{a}^*, \mathbf{p}^*)\}$ can be approximated with the current value of $f^i(\mathbf{a}^*, \mathbf{p}^*)$ or with a weighted time average $\sum_{j \geq 0} \alpha^j f^{i-j}(\mathbf{a}^*, \mathbf{p}^*)$, where α is a tuning parameter. The dynamic behavior of the iterative method, i.e the convergence speed and the residual error, depends on the choice of step-size parameters.

A. Convergence issues

Concerning the the global convergence of the proposed algorithm, the system utility in eq.(2) is nonconcave. Therefore, it appears that the proposed algorithm will not converge or will converge to only a locally optimal power/rate allocation, because it is based on solving the dual problem and yet the duality gap can be strictly positive. Note that spatial power allocation for downlink sumrate optimization is a non-convex problem and its solution is still open in the literature [14]. This fact has motivated our suboptimal approach leading to eq.(19) in the next Section, which precludes convergence to the global solution. However, simulations that we have done so far present: a stable behavior in stationary scenarios, a good trade-off complexity versus performance and good tracking capabilities. Moreover, the stochastic approximation procedure when applying the subgradient in order to solve the dual optimization problem, is quite well studied in the literature [15]. We omit its analysis for the sake of clarity.

V. ALLOCATION ALGORITHM

We are now deriving the allocation algorithm, i.e. the algorithm which provides at each time slot i , the optimal values of $\mathbf{a}^*, \mathbf{p}^*$ given λ, μ, ν . By keeping as main task the low complexity, we seek for a suboptimal solution. By using ergodic approximation² the Lagrangian can be maximized with respect to \mathbf{a}, \mathbf{p} , given the dual variables λ, μ, ν . The result can be expressed as follows. Let $\mathbf{u}_m = [u_{m,1}, \dots, u_{m,N_T}]$ with $u_{m,q} \in \{0, 1, \dots, K\}$ a vector of N_T user indexes (index $u_{m,q} = 0$ has the meaning of no user in position q), and $\mathbf{v} = [v_1, \dots, v_{N_T}]$ with $v_q \geq 0$ a vector of N_T power values. The solution is written for each frequency m as

$$\mathbf{u}_m^* = \arg \max_{\mathbf{u}_m} M(\mathbf{u}_m) \quad (13)$$

with

$$M(\mathbf{u}_m) = \sum_{q=1, u_{m,q} \neq 0}^{N_T} \begin{cases} (\nu u_{m,q} + w_{u_{m,q}}) F_{m,q}(\tilde{\mathbf{v}}) - \nu u_{m,q} R_{0k}/M - \lambda \tilde{v}_q & u_{m,q} \leq K_0 \\ (\mu u_{m,q} + w_{u_{m,q}} - \mu^T \phi) F_{m,q}(\tilde{\mathbf{v}}) - \lambda \tilde{v}_q & u_{m,q} > K_0 \end{cases} \quad (14)$$

where $F_{m,q}(\mathbf{v}) = \log_2(1 + \gamma_{u_{m,q},m,q}(\mathbf{v}))$,

$$\gamma_{u_{m,q},m,q}(\mathbf{v}) = \frac{v_q c_{u_{m,q},m,q}}{1 + \sum_{s \neq q, u_{m,s} \neq 0} v_s c_{u_{m,q},m,s}} \quad (15)$$

and $\tilde{\mathbf{v}} = \tilde{\mathbf{v}}(\mathbf{u}_m)$ is the solution of the set of N_T equations

$$\frac{d}{dv_q} \sum_{q=1, u_{m,q} \neq 0}^{N_T} \begin{cases} (\mu u_{m,q} + w_{u_{m,q}}) F_{m,q}(\mathbf{v}) - \lambda v_q & u_{m,q} \leq K_0 \\ (\mu u_{m,q} + w_{u_{m,q}} - \mu^T \phi) F_{m,q}(\mathbf{v}) - \lambda v_q & u_{m,q} > K_0 \end{cases} \quad (16)$$

Moreover,

$$a_{u_{m,q},m,q}^* = 1 \text{ if } u_{m,q}^* \neq 0, \quad a_{k,m,q}^* = 0 \text{ if } k \neq u_{m,q}^* \quad (17)$$

²This means that the statistical average is replaced by the time average

$$p_{u_{m,q},m,q}^* = \tilde{v}_q(\mathbf{u}_m^*) \quad (18)$$

Given a user set S , the optimization with respect to \mathbf{v} is convex; with respect to the set of scheduled users, however, an exhaustive search results in an excessive computational complexity. For this reason we distinguish between the scheduling phase (variable \mathbf{a}) and the power allocation (variable \mathbf{p}). The set (16) of N_T non linear equations is the power allocation problem for a given set of users that are spatially multiplexed on frequency m . This problem is usually analytically untractable. As suboptimal solution we will provide the water-filling solution evaluated by assuming constant uniform power allocation for the interfering beams:

$$\tilde{v}_q = \left[\frac{\nu u_{m,q} + w_{u_{m,q}}}{\lambda \log(2)} - \frac{V}{\gamma_{u_{m,q},m,q}(\mathbf{V})} \right]^+ \quad u_{m,q} \leq K_0$$

$$\tilde{v}_q = \left[\frac{\mu u_{m,q} + w_{u_{m,q}} - \mu^T \phi}{\lambda \log(2)} - \frac{V}{\gamma_{u_{m,q},m,q}(\mathbf{V})} \right]^+ \quad u_{m,q} > K_0 \quad (19)$$

where $\mathbf{V} = [V, V, \dots, V]$, and the power V is a parameters which estimates the power of interfering beams.

The equation (13) represents a discrete optimization problem which is referred as space-frequency allocation. This problem could be further simplified, by making it independent of power allocation, through the replacement of $\tilde{\mathbf{v}}$ with \mathbf{V} and $\gamma_{u_{m,q},m,q}(\tilde{\mathbf{v}})$ with $\gamma_{u_{m,q},m,q}(\mathbf{V})$. This simplified allocation can enjoy the possibility of a simplified SNIR feedback, which, on the other hand, does not allow the evaluation of the exact achievable rate to be used for adaptive modulation and coding.

Space-frequency allocation requires in general an exhaustive search in the set of all possible vectors \mathbf{u}_m . If we try to classify them we should say that there are $Q_m \in \{1, \dots, N_T\}$ active beams, there are $j_m \in \{1, \dots, N_T!/((N_T - Q_m)!Q_m!)\}$ dispositions of Q_m active beams, there are $h_m \in \{1, \dots, K!/(K - Q_m)!\}$ combinations of Q_m out of K users over each disposition of Q_m active beams. This huge search space can be reduced by using by using suboptimal algorithms. One of these algorithms is the one considered in [13] which can be applied when $\tilde{\mathbf{v}}$ is replaced with \mathbf{V} . A slightly different formulation which emphasizes the role of a scheduler as an entity that selects users according to a given metric is summarized here. It iterates the following operations until all the available MN_T space-frequency resources have been allocated:

- for each user k find the best resource (m_k, q_k) which maximizes $\gamma_{k,m,q}(\mathbf{V})$
- for each user k find the best other-user combination able to share frequency m_k with user k and the best beam disposition (including beam q_k) on frequency m_k .
- evaluate for the selected combination of users sharing frequency m_k the metric $M(\mathbf{u}_{m_k})$.
- (*) select for allocation the user combination (among the K evaluated before, one for each user) with the best metric and mark the just allocated frequency as unavailable.

The step marked with (*), as discussed later, can be executed by an entity denoted as scheduler, whereas the other steps are executed iteratively by an entity denoted as resource allocator³

VI. SYSTEM ARCHITECTURE

This section discusses the main architectural elements of a system that applies the proposed scheduling strategy to the transmission of

³These steps need only, at each iteration, a simple update which takes into account the new status of the available resources.

multi-user heterogeneous traffic flows in the downlink of a wireless system. In this discussion we refer to the scheme in Fig. 1 and the cross-layer aspects are emphasized.

A. Applications

We consider 3 different types of applications:

- Application generating VoIP traffic. The source has active periods where a constant bit-rate data flow is generated and silence periods. The traffic has stringent delay requirements and requires a minimum rate during active periods. Data transmitted has a timestamp indicating its deadline D which is used by the scheduler. Voice activity information is useful for scheduling to avoid assignment of unuseful resources. For this traffic, the minimum rate R_{0k} is equal to R_0 in the active periods and switched to 0 during silence period
- Application generating streaming video traffic. The source generates a variable rate data flow and the video packet have different roles and priorities inside the video stream. The traffic has delay requirements: according to playout delay at destination a deadline is fixed and packets arriving after deadline are discarded. Deadline D is used by the scheduler. The quality of the transmission is related to the amount of data arriving at the decoder within the deadline and to their importance in video reconstruction. Side information on the importance of video packet is sent to the queue manager. For this type of traffic the available rate is shared with fairness among users of the same class by tuning the parameters ϕ_k .
- Application generating FTP data traffic. In this case the traffic has not delay requirements. For this type of traffic the available rate is shared with fairness among users of the same class by tuning the parameters ϕ_k . The available rate in this case is lower than the rate assigned to streaming traffic.

The applications provides the lower layers with information on how to set parameters R_0 and ϕ_k , on the the maximum tolerable delay, on the deadline of packets and on the type of video packets.

B. Queue buffers

Data packets coming from the application are placed in a queue. There is a queue for each traffic flow. The queue manager provides the scheduler with the information on the time to deadline TD_k of the Head of Line (HOL) packet of each queue. The queue manager also implements a buffer management policy for each buffer. We consider here the following policies:

- Dropping of expired packets (DXP): packets with expired deadline are dropped to avoid waste of radio resource. This can be applied for VoIP traffic.
- Dropping based on packet priority and dependency (DPD): low priority packets with time to deadline below a given threshold are dropped from the buffer when the size of the queue is too large. This can be applied to streaming video traffic to prevent packet loss for late delivery in case of peaks in the source rate. This policy requires the knowledge of side information on the priority of the packets in the video stream.

C. Scheduler

This is the entity that decides which user or set of users is scheduled for transmission on part of the next available radio resources. This decision is a part of the iterative process for space-frequency allocation described in the previous Section and is based on the comparison of the metrics $M(\mathbf{u}_{m_k})$. These metrics depend on the weights w_k and on the resources that the resource allocator intends to

assign. The weights are dinamically set up by the scheduler to make its decision dependent on the status of the queues. The following two strategies are proposed for real time traffic:

$$w_k = 1 + be^{-aTD_k/\tau_{max}} \quad (20)$$

and

$$w_k = 1 + \beta_k e^{-aTD_k/\tau_{max}} \quad (21)$$

where $\beta_k = b(\nu_k + 1)$, if $k \leq K_0$, and $\beta_k = b\mu_k$, if $k > K_0$. Parameters a and b are used to suitably shape the functions of the time to deadline TD_k of the HOL packet in the queue and τ_{max} is the maximum tolerable delay of the application. For both strategies the weight w_k increases exponentially as the TD_k decreases; in the second strategy the increase is proportional to dual variable μ_k or ν_k to improve fairness. The scheduler, by using the parameters R_{0k} and ϕ_k and the information on rate allocated to user provided by the resource allocator, keeps track of rate constraints and updates the values of dual variables μ_k, ν_k . It also sends to resource allocator the updated values of w_k and μ_k, ν_k . This entity does not interact with the physical layer.

D. Resource allocator

This is the entity which decides which resources (space, frequency and power) can be assigned to a given set of users. This decision is a part of the iterative process for space-frequency allocation described in the previous Section and is based on the CSI provided by the physical layer, and the information provided by the scheduler. It evaluates for the scheduler the metrics $M(\mathbf{u}_{m_k})$ and the rate achievable with the assigned resources. When a set of users is scheduled for transmission with a set of assigned resources, the resource allocator sends to physical layer information on the power to use and on the modulation and coding format suitable to realize the assigned rate. The resource allocator is air-interface aware, but does not interact with the queues. The resource allocator also keeps track of power constraint and updates the value of λ .

E. Physical layer

Performs the transmission of scheduled data, according to the resources assigned by the resource allocator. It also handles CSI feedback, which is collected and periodically sent to resource allocator.

VII. NUMERICAL RESULTS

Results are obtained for a scenario which incorporates some characteristic aspects of practical applications in next generation wireless systems. We are considering here a single cell of the downlink of an OFDMA wireless system with $M=128$ subcarriers working on a bandwidth of 1.25 Mhz. Base station is equipped with multiple antennas. The system is TDD and it is assumed that 2/5 of frame interval is used for downlink transmission. The CSI coming from users is updated every 10ms and for rate assignment a signal-to-noise-raio gap of 3dB is adopted. The users have a position which is uniformly distributed in circular area of radius 500m. The simulation time is 100s.

Channel model includes path loss, correlated shadowing (not present in the second option for user distributions) and time and frequency correlated fast fading. Path loss is modeled as a function of distance as $L(dB) = k_0 + k_1 \log(d)$ ($k_1=40$, $k_2=15.2$ for results). Shadowing is superimposed to path-loss, with classical lognormal model (std. deviation: 6 dB) and exponential correlation in space (correlation distance equal to 20m). Fast fading on each link of the MIMO broadcast channel is complex Gaussian, independent across antennas and is modeled according to a 3GPP Pedestrian model. This

model has a finite number of complex multipath components with fixed delay (delay spread of 2.3 microseconds) and power (average normalized to 1). Time correlation is obtained according to a Jakes' model with given Doppler bandwidth (6 Hz in the results). At the base station orthogonal beamforming is adopted, where beam vectors change randomly at each frame. In the simulated system the total average power constraint is fixed to 1W. To obtain the results we also assume that channel variations in time due to Doppler effects have a negligible impact on the feedback quality. The effects of outdated feedback would require a proper investigation which is out of the scope of the paper.

The first two Tables I and II have the aim of investigating the behavior of the scheduling algorithm with respect to different design options for the weights w_k , the step size δ_ν and the update of variables ν_k for voice users, the use of buffer managements strategies. The performance figures considered are the average rate assigned to users (efficiency at the physical layer), the achieved PSNR for video transmission and the drop-rate for VoIP transmission (quality at the application). The fixed parameters are: $R_{0k}=64$ kbit/s, $\phi_k=1/12$, $\alpha=0.75$, $\delta_\lambda=0.01$, $\delta_\mu=0.00005$. The first remark is the substantial fairness achieved among the different video users in spite of the very different channel conditions. We also found the following elements. First, the best way to set up weights w_k is as in eq. (21) which exploits information on packet deadline coming from the queues to prioritize in the short term packets approaching deadlines. This is important for variable rate traffic flows. Second, the periodical update of variables ν_k during the silence periods of voice transmission is very important to avoid data dropping⁴ in the first part of voice talk periods where subgradient algorithms needs some time to converge to the new values of ν_k . The update of variables ν_k according to eqs.(10)-(12) requires a virtual assignment of rate which is not used in silence periods. Third, the use of suitable packet dropping strategies at buffer level is important to improve achieved quality at the application level: dropping packets that will be useless if transmitted avoids waste of resource; a content-aware dropping of selected packets in a video stream improves the quality at the decoder.

Figures 2 and 3 explore the dynamic behavior of the scheduling algorithm for the system configuration corresponding to the last line of Table II. Fig. 2 shows the behavior of the power which the resource allocator would assign according to the constraint on the average total power as in the problem (2). Since the variations of instantaneous power around the average are reasonably limited, it is possible to apply instantaneous power scaling after power assignments to satisfy the constraint P even for the instantaneous power. When needed, and this does not destroy the efficiency of the scheduling algorithm. Fig. 3 shows, for one video user, the behavior of variable μ_k and of weight w_k according to the state of the channel and to the status of the queue (which is illustrated through the graph of the normalized time-to-deadline of the HOL packet). It is interesting to note that w_k reacts to the state of the queue in favor of situations where packets approach the deadline, whereas μ_k follows the state of the entire system with its slow channel variations⁵, trying to accommodate average rate constraints.

Figure 4 has the aim of showing the capability of the scheduler to allocate physical layer resources according to rate constraints imposed in eq. (2). In the system there are 4 different classes of users with 4 different rate requirements (1 class with constant rate constraint,

⁴the residual dropping rate of one voice user in this set of simulated results is due to a very deep shadowing lasting for nearly 10s.

⁵fast fading is handled by the space-frequency scheduler

TABLE I
SCENARIO WITH 12 VIDEO USERS. RATE ASSIGNED BY THE SCHEDULER AND PSNR FOR DIFFERENT CHOICES OF WEIGHTS w_k . THE RESULTS IN THE LOWER PART OF THE TABLE REFER TO THE USE OF DPD IN THE BUFFER.

param.	Tot. rate [kbit/s]	Rate [kbit/s]	av. PSNR [dB]	PSNR range [dB]
$w_k=1$,	2035	160-175	31.1	2.5
(20), $b=2$, $a=2$	2240	170-210	33.3	6
(21), $b=4$, $a=2$	2069	167-177	32.3	2
$w_k=1$,	2030	160-175	33.3	1.8
(20), $b=2$, $a=2$	2175	168-198	34.8	3.3
(21), $b=4$, $a=2$	2067	165-178	34.1	1.8

TABLE II
SCENARIO WITH 12 VIDEO USERS AND 3 VOICE USERS. RATE ASSIGNED BY THE SCHEDULER AND QUALITY (PSNR FOR VIDEO AND DROP-RATE FOR VOICE) FOR DIFFERENT CHOICES OF WEIGHTS w_k , PARAMETERS b AND δ_ν , USE OF THE UPDATE (UP) OF DUAL VARIABLES ν_k IN SILENCE PERIODS OF VOICE TRANSMISSION.

param.	Rate video [kbit/s]	Rate voice [kbit/s]	PSNR [dB]	drop-rate [%]
$w_k=1$,				
$\delta_\nu=0.0002$	155-171	40, 63, 75	31.4-34.4	64, 40, 19
$w_k=1$,				
$\delta_\nu=0.001$	145-168	58, 60, 70	31.4-34.4	48, 27, 18
(20), $b=2$,				
$\delta_\nu=0.0002$	160-181	44, 61, 77	31.9-35.3	53, 20, 16
(21), $b=4$,				
$\delta_\nu=0.0002$	153-170	48, 61, 75	32.8-34.8	42, 14, 2.6
(21), $b=4$,				
$\delta_\nu=0.001$	149-165	62, 62, 68	32.6-34.7	25, 5.1, 4.1
(21), $b=4$,				
$\delta_\nu=0.0002$,	152-165	73, 87, 102	32.7-34.9	34, 1.6, 0.1
UP: 4 frames				
(21), $b=4$,				
$\delta_\nu=0.001$	146-159	79, 88, 93	32.4-34.4	19, 0, 0
UP: 4 frames				
(21), $b=4$, 8,				
$\delta_\nu=0.001$	143-158	80, 81, 95	32.5-34.2	17, 0.4, 0
UP: 10 frames				

3 classes with proportional rate constraints). In spite of different channel conditions the constraints are respected with fairness inside each class.

Finally, the last Table investigate how the assigned rate and quality of service indicators change by increasing the number of user in the system. In this case, the users are assumed to be placed at the same distance $d=250$ m from the base station and do not experience shadowing effects. It is shown that when the user load increases the system is able to preserve the quality of voice users, whereas the quality perceived by the FTP user and the video user decreases, without losing fairness.

VIII. CONCLUSIONS

In this paper we presented a cross-layer multiuser scheduling strategy for a MIMO-OFDMA system which tries to maximize sum-rate whilst supporting with fairness different traffic classes with different constraints. It exploits temporal diversity by means of ergodic maximization, it allows low complexity on-line implementation and uses a reduced feedback thanks to opportunistic beamforming. We have shown in the numerical results that the algorithm is able to allocate channel resources according to the constraints of each different traffic class and with fairness inside each class, even in presence of different channels conditions and with different network loads.

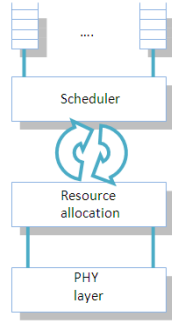


Fig. 1. Simplified architecture of the system that implements the proposed

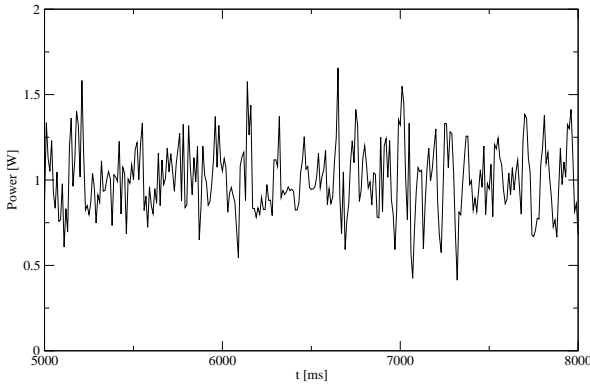


Fig. 2. Dynamic behavior of the total power assigned by the scheduler without power scaling. Scenario with 12 video users and 3 voice users, parameters as in the last line of Table II.

REFERENCES

- [1] J. Brehmer, W. Utschick, "Nonconcave Utility Maximisation in the MIMO Broadcast Channel," *Eurasip JASP*, vol. 2009.
- [2] M. Haardt, V. Stankovic, G. Del Galdo, "Efficient Multi-user MIMO Downlink Precoding and Scheduling," *CAMSAP 05*, Dec. 2005.
- [3] Sriram Viswanath, Nihar Jindal, and Andrea Goldsmith, "On the Capacity of Multiple Input Multiple Output Broadcast Channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 15701580, Apr. 2005.
- [4] W. Yu, "Uplink-Downlink Duality Via Minimax Duality," *IEEE Trans. on IT*, February 2006.
- [5] E. Jorswieck, A. Sezgin, B. Ottersten, A. Paulraj, "Feedback reduction in uplink MIMO OFDM Systems by Chunk Optimization," *Eurasip JASP*, Vol. 2008.
- [6] N. Wei, A. Pokhariyal, T. B. Sorensen, T. E. Kolding, P.E. Mogensen, *IEEE Journal on Sel. Areas in Comm.*, vol.26, no. 6, Aug. 2008.

TABLE III

QUALITY FIGURES AS A FUNCTION OF THE NUMBER OF VIDEO USERS IN A SCENARIO WITH 3 OTHERS VOICE USERS AND 1 FTP USER DOWNLOADING 2MB OF DATA.

users	Tot.rate [kbit/s]	PSNR [dB]	drop-rate [%]	max. delay [s]
10	184	36.3	0	5.33
12	198	35.9	0	9.17
14	209	35.5	0	14.85
16	217	34.9	0	20.64
18	220	34.1	0	32.06
20	223	33.5	0.5	41.22
20 (d=300m)	191	32.5	0	50.57
20 (d=350m)	160	30.8	0	66.07

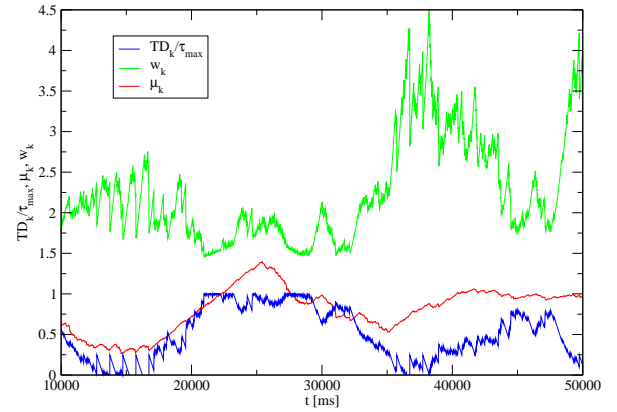


Fig. 3. Dynamic behavior of normalized time-to-deadline of the HOL packet in the queue, the dual variable u_k and weight w_k for one video user. Scenario

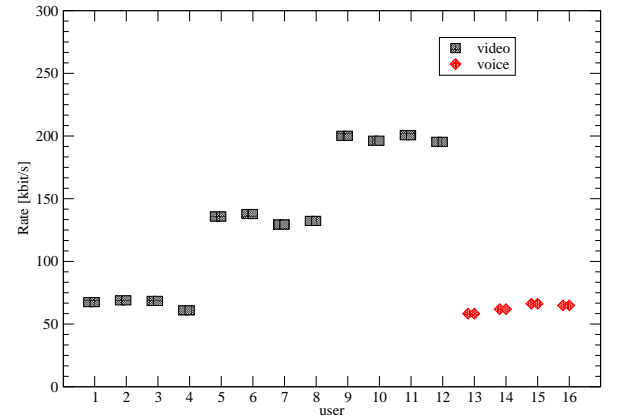


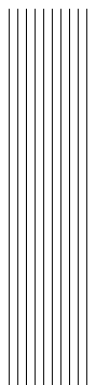
Fig. 4. The different rates assigned by the scheduler in a scenario with 3 different classes of 4 video users ($\phi_k \in \{0.5\phi, \phi, 1.5\phi\}$, with $\phi = 1/12$) and 4 voice users, parameters as in the last line of Table II.

- [7] M. Codrenau, M. Juntti, M. Latva-Aho, "Low-complexity iterative algorithm for finding the MIMO-OFDM broadcast channel sum capacity," *IEEE Trans. on Comm.*, vol. 55, no. 1, Jan 2007.
- [8] Issam Toufik, Marios Kountouris, "Power allocation and feedback reduction for MIMO-OFDMA opportunistic Beamforming," *VTC Spring 2006*: pages 2568-2572
- [9] G. Liu; J. Zhang; F. Jiang; W. Wang, "Joint Spatial and Frequency Proportional Fairness Scheduling for MIMO OFDMA Downlink," *IEEE WiCom 2007*, 21-25 Sept. 2007 Page(s):491 - 494
- [10] Q. Liu, X. Wang, and G. B. Giannakis, "A cross-layer scheduling algorithm with QoS support in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 55, pp. 839847, May 2006.
- [11] V. Corvino, V. Tralli, and R. Verdone, "Cross layer radio resource allocation for multi carrier air interfaces, in multi-cell multi-user environments," *IEEE Trans. Veh. Technol.*, to appear
- [12] M. Sharif and B. Hassibi, "On the Capacity of MIMO Broadcast Channel with Partial Side Information," *IEEE Trans. on Inform. Theory*, vol. 51, no. 2, pp. 506-522, Feb 2005.
- [13] Ana I. Perez-Neira, Pol Henarejos, Velio Tralli, Miguel A. Lagunas, "A low complexity space-frequency multiuser resource allocation algorithm", in *Proceedings of WSA 2009*, Berlin, Germany, February 16-18, 2009
- [14] Mung Chiang, Chee Wei Tan, Daniel P. Palomar, Daniel O'Neill, and David Julian, "Power Control By Geometric Programming," *IEEE Trans on Wireless Comm.*, VOL. 6, NO. 7, JULY 2007
- [15] Y. Ermoliev, R. Wets, "Numerical Techniques for Stochastic Optimization," Springer-Verlag, 1988.

Índex de figures

1	Interpretació visual del <i>water-filling</i> . La potència assignada és l'àrea en vermell per sota d'un llindar.	xvi
1.1	Esquema del projecte.	3
2.1	Entropia de dos successos amb probabilitats $p(X_1), p(X_2) = 1 - p(X_1)$	7
2.2	Sistema de comunicació	9
2.3	Recorregut d'un senyal amb múltiples camins.	13
2.4	Nivell de potència rebut en funció de la distància seguint el model anterior . .	14
2.5	Planificació freqüencial d'OFDMA.	16
2.6	Descomposició SVD.	20
3.1	Regió de capacitat de MAC.	25
3.2	Regió de capacitat de BC.	28
3.3	Descomposició RQ.	32
3.4	Aplicació RQ.	33
3.5	Estratègia MOB.	34
3.6	Comparació de les estratègies TDMA, MOB i DPC per $K = 2$, $K = 100$ i $K = \infty$ amb $N_T = 2$	35
3.7	Regió QoS.	36
4.1	Diagrama de blocs de tot el procés	42
5.1	Velocitat mitja normalitzada amb 5 usuaris.	54
5.2	Evolució instantània de la λ	54
5.3	Evolució amitjanada de la μ	54
5.4	Evolució amitjanada de la potència.	55
5.5	Velocitats instantànies per usuari.	55
5.6	Evolució instantània de la μ	56
5.7	Evolució instantània de la potència.	56
5.8	Velocitat mitja per usuari amb canvi d'usuaris.	57
5.9	Evolució amitjanada de la μ amb canvi d'usuaris.	57
5.10	Evolució instantània de la λ amb canvi d'usuaris.	57
5.11	Evolució amitjanada de la potència amb canvi d'usuaris.	58

5.12	Velocitat instantània per usuari amb canvi d'usuaris.	59
5.13	Evolució instantània de la μ amb canvi d'usuaris.	59
5.14	Evolució instantània de la potència amb canvi d'usuaris.	59
5.15	Regió de velocitat per dos usuaris amb múltiples implementacions.	60
5.16	Consum de temps de l'algoritme en diferents configuracions i usuaris.	62
5.17	<i>Sumrate</i> en funció del nombre d'antenes per $K = 8$ i $K = 32$ usuaris i múltiples configuracions.	62
5.18	<i>Sumrate</i> per 10 usuaris a 250m de distància.	64
5.19	Evolució de la potència total per 10 usuaris a 250m de distància.	64
5.20	<i>Sumrate</i> per 20 usuaris a 250m de distància.	65
A.1	Funcions de densitat de probabilitat Rayleigh per diferents variàncies.	74
A.2	Funcions de densitat de probabilitat Rice per diferents valors de α	75
A.3	Funcions de distribució de probabilitat lognormal per diferents variàncies.	75
A.4	Conseqüència del <i>time-sharing</i>	78



Índex de taules

4.1	Algoritme pas a pas amb la complexitat	49
-----	--	----



Bibliografia

- [ADS16] *ETSI TS 101 388* “Transmission and Multiplexing (TM); Access transmission systems on metallic access cables; Asymmetric Digital Subscriber Line (ADSL) - European specific requirements”, ETSI, 2002.
- [BGI32] A. Ribeiro and G. B. Giannakis. “Layer Separability of Wireless Networks”. *IEEE Information Sciences and Systems, 2008. CISS 2008*, març 2008.
- [COS09] Max H. M. Costa. “Writing on Dirty Paper”. *IEEE Transactions on Information Theory*, vol. *IT-29*, maig 1983.
- [DAB14] *ETSI EN 300 401* “Radio Broadcasting Systems; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers”, ETSI, 2001.
- [DVB20] *ETSI EN 301 192* “Digital Video Broadcasting (DVB); DVB specification for data broadcasting”, ETSI, 2004.
- [GOL10] N. Jindal, S. Vishwanath and A. Goldsmith. “On the Duality of Gaussian Multiple-Access and Broadcast Channels”. *IEEE Transactions on Information Theory*, vol. *50*, no. *5*, maig 2004.
- [GOL12] N. Jindal, W. Rhee, S. Vishwanath, S.A. Jafar and A. Goldsmith. “Sum Power Iterative Water-filling for Multi-Antenna Gaussian Broadcast Channels”. *IEEE Transactions on Information Theory*, vol. *51*, abril 2005.
- [GOL22] S. Vishwanath, N. Jindal and A. Goldsmith. “Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels”. *IEEE Transactions on Information Theory*, vol. *49*, no. *10*, octubre 2003.
- [GOL25] S. Vishwanath, N. Jindal and A. Goldsmith. “On the Capacity of Multiple Input Multiple Output Broadcast Channels”. *IEEE Transactions on Information Theory*, vol. *51*, abril 2005.
- [HAA05] Bernd Bandemer, Martin Haardt and Samuli Visuri. “Linear MSE Multi-user MIMO Downlink Precoding for Users with Multiple Antennas”. *17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06)*, 2006.

- [HAA06] Quentin H. Spencer and Martin Haardt. "Capacity and Downlink Transmission Algorithms for a Multi-user MIMO Channel". *IEEE*, 2002.
- [HAA13] Q.H. Spencer, A.L. Swindlehurst and M. Haardt. "Zero-Forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels". *IEEE Transactions on Signal Processing*, vol. 52, febrer 2004.
- [HAA24] M. Haardt, V. Stankovic and G. del Galdo. "Efficient Multiuser MIMO Downlink Precoding and Scheduling". *CAMSAP 05*, desembre 2005.
- [HAN33] M. Chiang, S. Zhang, P. Hande. "Distributed rate allocation for inelastic flows: optimization frameworks, optimality conditions, and optimal algorithms". *IEEE JSAC*, vol. 23, no.1, gener 2005.
- [HAS34] M.Sharif, B.Hassibi. "On the capacity of Broadcast Channels with Partial Side Information". *IEEE Transactions on Information Theory*, febrer 2005.
- [IEE15] *IEEE. Std 802.11a*: "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications", 1999.
- [JUN11] M. Codrenau, M. Juntti and M. Latva-Aho. "Low-Complexity Iterative Algorithm for Finding the MIMO-OFDM Broadcast Channel Sum Capacity". *IEEE Transactions on Communications*, vol. 55, no. 1, gener 2007.
- [KOB08] M. Kobayashi and G. Caire. "An Iterative Water-Filling Algorithm for Maximum Weighted Sum-Rate of Gaussian MIMO-BC". *IEEE JSAC*, vol. 24, agost 2006.
- [KOU07] Issam Toufik and Marios Kontouris. "Power allocation and feedback reduction for MIMO-OFDMA opportunistic Beamforming". *VTC Spring 2006: 2568-2572*, 2006.
- [MAX01] J. C. Maxwell. "On Physical Lines of Force", març 1861.
- [ONE36] M. Chiang, C. W. Tan, D. P. Palomar, D. O'Neill and David Julian. "Power Control By Geometric Programming". *IEEE Transactions on Wireless Communication*, vol. 6, no. 7, juliol 2007.
- [OTT27] E. Jorswieck, A. Sezgin, B. Ottersten and A. Paulraj. "Feedback reduction in uplink OFDM Systems by Chunk Optimization". *Eurasip JASP*, 2008.
- [PER19] A. Pérez-Neira, P. Henarejos, V. Tralli and M. A. Lagunas. "A Low Complexity Space-Frequency Multiuser Scheduling Algorithm". *International ITG Workshop on Smart Antennas*, març 2009.
- [SHA02] C. E. Shannon. "A Mathematical Theory of Communication", 1948.
- [SHM21] H. Weingarten, Y. Steinberg and S. Shamai. "The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel". *IEEE Transactions on Information Theory*, vol. 52, no. 9, setembre 2006.
- [TAE04] Tae-Sung Kang and Hyung-Myung Kim. "Optimal Beam Subset and User Selection for Orthogonal Random Beamforming". *IEEE Communications letters*, vol. 12, no. 9, setembre 2008.

- [TRA18] V. Tralli, A. Pérez-Neira and P. Henarejos. “A cross-layer multiuser scheduling strategy for heterogeneous traffic in the downlink of a MIMO-OFDMA systems”. *IEEE Globecom 2009 Wireless Communications Symposium*, 2009. Pendent d’acceptació.
- [UTS23] J. Brehmer and W. Utschick. “Nonconcave Utility Maximization in the MIMO Broadcast Channel”. *Eurasip JASP*, 2009.
- [VTL17] P. Viswanath, D. N. C. Tse and R. Laroia. “Opportunistic beamforming using dumb antennas”. *IEEE Transactions on Information Theory*, 48:1277-1294, juny 2002.
- [WAN28] G. Liu, J. Zhang, F. Jiang and W. Wang. “Joint Spatial and Frequency Proportional Fairness Scheduling for MIMO OFDMA Downlink”. *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on vol. issue 21-25*, setembre 2007.
- [WIF30] *IEEE Std 802.11g-2003* “Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications”, IEEE, juny 2003.
- [WIM29] *IEEE Std 802.16e-2005* “IEEE Standard for Local and metropolitan area networks”, IEEE, febrer 2005.
- [WON35] I. Wong, B. Evans. “Optimal OFDMA resource allocation with linear complexity to maximize ergodic weighted sum capacity”. *ICASSP 07, Hawaii*, 2007.
- [WYU26] W. Yu. “Uplink-Downlink Duality Via Minimax Duality”. *IEEE Transactions on Information Theory*, febrer 2006.
- [WYU31] W. Yu, R. Lui and R. Cendrillon. “Dual Optimization Methods for Multiuser Orthogonal Frequency-Division Multiplex Systems”. *IEEE Global Telecommunications Conference, 2004. GLOBECOM '04*, desembre 2004.
- [ZIM03] H. Zimmermann. “OSI Reference Model — The ISO Model of Architecture for Open Systems Interconnection”. *IEEE Transactions on Communications*, vol. Com-28, no. 4, abril 1980.



Crèdits

A continuació es detallen els crèdits corresponents a les figures utilitzades.

- Figura 2.4: Grup Comunicacions Mòbils TSC-UPC.
- Figura 3.5: Nizar Zorba.
- Figura 3.6: Nizar Zorba.
- Figura 2.3: Dr. Dharma, P. Agawald and Qing-An Zeng.
- Figura A.4: W. Yu.